DOCUMENT RESUME

ED 468 690 CS 511 387

AUTHOR	Taylor, Barbara M.; Pearson, P. David
TITLE	The CIERA School Change Project: Supporting Schools as They Implement Home-Grown Reading Reform. CIERA Report.
INSTITUTION	Center for the Improvement of Early Reading Achievement, Ann Arbor.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CIERA-2-016
PUB DATE	2002-06-13
NOTE	56p.
CONTRACT	R305R70004
AVAILABLE FROM	CIERA/University of Michigan, 610 E. University Ave., 1600 SEB, Ann Arbor, MI 48109-1259. Tel: 734-647-6940; Fax: 734-763-1229; Web site: http://www.ciera.org.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	*Educational Change; Elementary Education; Instructional Effectiveness; Phonics; Poverty; *Reading Achievement; Reading Fluency; *Reading Instruction; *Teaching Methods; *Writing Instruction
IDENTIFIERS	Phonemic Awareness

ABSTRACT

This work investigates the relationship between the programmatic and classroom instructional factors of schools and students' reading and writing achievement in 14 high-poverty schools. The study finds that schools which rated higher on a scale of collaborative leadership showed greater student growth in reading fluency and writing. Classroom observations revealed that telling students information and engaging them in recitation were negatively related to reading growth in grades two through six, whereas active responding was positively related to reading growth in grades four through six. In grades one through six, asking students higher-level questions about text after reading was positively related to student growth. A high level of phonics instruction was negatively related to reading growth in kindergarten and grades two and three, whereas a high level of phonemic awareness instruction was positively related to growth in phonemic segmentation and blending ability in kindergarten. In grades four through six, the practice of coaching children in the use of word recognition strategies during reading was positively related to reading growth. Engaging students in small group instruction was positively related to reading growth in kindergarten and grade one. An appendix contains the rubric for rating interview responses. (Contains 25 tables, 5 notes, and 40 references.) (PM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 468 690

The CIERA School Change Project: Supporting Schools as They Implement Home-Grown Reading Reform

CIERA REPORT #2-016

Barbara M. Taylor
Debra Peterson
Michael C. Rodriguez
University of Minnesota

P. David Pearson
University of California-Berkeley

CIERA Inquiry 2: Home and School
Which school-level and classroom-level variables maximize students' growth in reading and writing? Can schools' reform efforts increase their use of effective practices to enhance students' reading growth?

Set within a grass-roots effort to implement reading program reform, this work investigates the relationship between the programmatic and classroom instructional factors of schools and students' reading and writing achievement. Examining these relationships across 14 high-poverty schools, we found that schools which rated higher on a scale of collaborative leadership showed greater student growth in reading fluency and writing. From our classroom observations, we learned that telling students information and engaging them in recitation were negatively related to reading growth in grades 2–6, whereas active responding was positively related to reading growth in grades 4–6. In grades 1–6, asking higher-level questions about text after reading was positively related to student' growth in reading and/or writing. A high level of phonics instruction was negatively related to emergent literacy or reading growth in kindergarten and grades 2–3, but not in

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

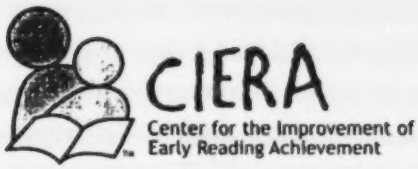
CIERA

Center for the Improvement of Early Reading Achievement
University of Michigan – Ann Arbor

June 13, 2002

BEST COPY AVAILABLE

CS 511 387



grade 1, whereas a high level of phonemic awareness instruction was positively related to growth in phonemic segmentation and blending ability in kindergarten. In grades 4–6, the practice of coaching children in the use of word recognition strategies during reading was positively related to reading growth. Engaging students in small group instruction was positively related to reading growth in kindergarten and grade 1, whereas whole-class or large-group instruction was positively related to reading growth in grades 4–6. Results are discussed within the framework of research on effective schools and teachers.

University of Michigan School of Education

610 E University Ave., Rm 1600 SEB
Ann Arbor, MI 48109-1259

734.647.6940 voice
734.615.4858 fax
ciera@umich.edu

www.ciera.org

©2002 Center for the Improvement of Early Reading Achievement.

This research was supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. However, the comments do not necessarily represent the positions or policies of the National Institute of Student Achievement, Curriculum, and Assessment or the National Institute on Early Childhood Development, or the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

The CIERA School Change Project: Supporting Schools as They Implement Home-Grown Reading Reform

Barbara M. Taylor, University of Minnesota
P. David Pearson, University of California-Berkeley
Debra Peterson, University of Minnesota
Michael C. Rodriguez, University of Minnesota

We know a great deal about what schools and teachers can do to promote reading success in the elementary grades. We also possess a great deal of knowledge about school change, and the importance of professional development. However, we are challenged by our apparent inability to put our knowledge to work. Even though we continue to learn more about effective schools, effective instruction, and effective change efforts, we seem hard-pressed to integrate and apply this knowledge in ways that positively impact the thousands of schools which are struggling to teach all children to read.

Research on Effective Schools

In the past, numerous studies of high-performing high-poverty schools have pointed to important building-level factors that must be in place in order for all children to achieve at high levels in reading. Emphasizing outcomes in reading achievement, Hoffman (1991) summarized the research on effective schools from the 1970s and early 1980s (e.g., Venezky & Winfield, 1979; Weber, 1971; Wilder, 1977). He discussed eight recurring attributes of effective schools:

1. a clear school mission;
2. effective instructional leadership and practices;
3. high expectations;

4. a safe, orderly, and positive environment;
5. ongoing curriculum improvement;
6. maximum use of instructional time;
7. frequent monitoring of student progress; and
8. positive home-school relationships.

In recent years, we have seen a revival of effective schools research, most likely due to widespread national concerns about student reading achievement. Taylor, Pressley, and Pearson (2002) summarized findings from five large-scale research studies on effective, high-poverty elementary schools, which were published between 1997 and 1999 (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein, Johnson, & Ragland, 1997; Puma, Karweit, Price, Ricciuiti, Thompson, & Vaden-Kiernan, 1997; Taylor, Pearson, Clark, & Walpole, 2000). The six recurring themes that emerge from these five studies both support and extend the earlier research on effective schools.

Putting the students first to improve student learning. In four of these studies (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein et al., 1997; Taylor et al., 2000), improved student learning was cited as the schools' overriding priority. Also, schools reported a collective sense of responsibility for school improvement. Teachers, parents, the principal, and other school staff members worked as a team to achieve their goal of substantially improved student learning and achievement.

Strong building leadership. Three of the studies (Designs for Change, 1998; Lein et al., 1997; Puma et al., 1997) documented the importance of strong building leadership. The principal may have worked to redirect people's time and energy, to develop a collective sense of responsibility for school improvement, to secure resources and training, to provide opportunities for collaboration, to create additional time for instruction, and to help the school staff persist in spite of difficulties.

Strong teacher collaboration. In addition to, or perhaps because of, strong leadership, strong staff collaboration was highlighted in four of the studies (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein et al., 1997; Taylor et al., 2000). Teachers planned and taught together, with a focus on how to best meet students' needs. They reported a strong sense of building communication, talking and working across, as well as within, grades, which contributed to better understanding of one another's curricula and expectations.

Focus on professional development and innovation. Four of the studies (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein et al., 1997; Taylor et al., 2000), stressed ongoing professional development and the implementation of new research-based practices. Many of the successful schools in these studies, emphasized a type of sustained professional development in which teachers learned together within a building and collaborated to improve their instruction.

Consistent use of student performance data to improve learning. Four of the studies (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein et

al., 1997; Taylor et al., 2000) found that teachers in effective schools systematically shared student assessment data, usually on curriculum-embedded measures, as a part of the process of making instructional decisions to improve pupil performance. Teachers also worked together to carefully align instruction to standards and state or district assessments.

Strong links to parents. All five studies (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein et al., 1997; Puma et al., 1997; Taylor et al., 2000) reported strong efforts within schools to reach out to parents. Schools worked to win the confidence of parents and then built effective partnerships with them in order to support student achievement. Parents were treated as valued members of the school community. Schools also reported a positive school climate, good relations with the community, and high levels of parental support.

Research on effective school reform and professional development. Research on effective school reform and teacher professional development is consistent with the research on effective schools in general, in that it stresses the importance of teachers learning and changing together over an extended period of time, as they reflect on their practice and implement new teaching strategies (Fullan, 2000; Fullan & Hargreaves, 1996; Louis & Kruse, 1995; Richardson & Placier, in press.) In successful schools, which typically operate as strong professional learning communities, teachers systematically study student assessment data, relate the data to their instruction, and work with others to refine their teaching practices (Fullan, 2000). Reflective dialogue, deprivatization of practice, and collaborative efforts all enhance shared understandings and strengthen relationships within a school (Louis & Kruse, 1995).

Research on Effective Teachers of Reading

The knowledge base for effective teaching, especially teaching reading in the elementary grades, is equally as strong. In a recent NEA research report, Taylor, Pressley, and Pearson (2002) summarize this research, noting several distinct historical waves of work. From the process-product research of the 1960s and 1970s (Brophy, 1973; Dunkin & Biddle, 1974; Flanders, 1970; Soar & Soar, 1979; and Stallings & Kaskowitz, 1974) we learned that more effective teachers maintained an academic focus, kept a high incidence of pupils on task, and provided direct instruction. Effective direct instruction included making learning goals clear, asking students questions as part of monitoring their understanding of what was being covered, and providing feedback to students about their academic progress. Effective classrooms were found to be warm, democratic, and cooperative, with more teacher instruction devoted to weaker students, who were also given more time to complete tasks.

A second wave of research on teaching reading, which began with the work of Duffy and Roehler in the 1980s, taught us about the cognitive processes used by outstanding teachers. More effective teachers engaged in modeling and explanation to teach students strategies for decoding words and under-

standing texts. Knapp and associates (Knapp, 1995) found that effective teachers stressed higher-level thinking skills more than lower-level skills. Continuing in this tradition, Taylor et al. (2000) found that accomplished primary grade teachers provided more small-group than whole-group instruction, had high pupil engagement, had a preferred teaching style of coaching as opposed to telling, and engaged students in more higher-level thinking related to reading than other teachers.

In the most recent wave of research, Pressley and his colleagues (Pressley, Wharton-McDonald, Allington, Block, Morrow, Tracey, Baker, Brooks, Cronin, Nelson, & Woo, 2001) have focused our attention on the characteristics of teachers nominated as exemplary in practice by their peers and supervisors. These researchers found that effective primary grade teachers did provide a balanced literacy program: they taught skills and got their students actively engaged in a great deal of actual reading and writing. They also encouraged students to self-regulate their use of strategies. Interestingly, the National Reading Panel Report (2000) implicated balanced literacy instruction in its conclusion that instructional attention to systematic phonics, phonemic awareness, fluency, and comprehension strategies was important to a complete reading program. (pp. 2–89).

In short, we have learned different, but complementary, lessons about the teaching practices of excellent elementary literacy teachers from the last four decades of research on effective teaching. The overall picture is consistent with the earlier process-product research to some extent, especially with regard to engagement, but goes beyond it in ways consistent with Duffy, Roehler, et al.'s (1987) direct explanation approach and Knapp and associates' (1995) emphasis on higher-order literacy instruction (i.e., instruction which emphasizes comprehension and communication). Excellent elementary literacy teachers balance skills instruction with more holistic teaching (Pressley, 1998). In the best classrooms, students are engaged much of the time in reading and writing, with the teacher monitoring student progress, encouraging continuous improvement and growth, and providing scaffolded instruction to help students improve their use of various strategies.

Objectives of the Current Project

Amidst pressure for schools to adopt off-the-shelf reform programs as a way of improving student achievement (Herman, 1999), it is interesting to note that, by and large, the schools in the studies summarized by Taylor, Pressley, and Pearson (2002) did not necessarily view packaged reforms as the key ingredient for improving student achievement (Charles A. Dana Center, 1999; Designs for Change, 1998; Taylor et al., 2000).[1] The common denominators seem to be commitment and hard work focused on research-based practices at both the classroom level and the school level.

---

1. This is not to say that prepackaged programs are less desirable or successful than homegrown reform programs. In fact, the majority of current research studies on reading programs in school reform efforts is focused on these prepackaged programs (e.g., Herman, 1999).

The overall objective of this project is to test the efficacy of a school reform framework which was designed to be used by elementary schools, in order to develop local reading programs that would improve students' reading achievement. The study is guided by two fundamental questions:

1. Will a research-based, action-oriented, Internet-supported change framework designed to promote a grass-roots reading program reform produce robust changes in: a) school-wide approaches to delivering reading instruction; b) classroom teaching practices; and c) student learning and reading achievement?

2. Across schools, what classroom practices and schoolwide efforts are most effective in improving the teaching of reading and increasing students' reading achievement?

In our attempt to answer these questions, we did not, nor do we think that we can or should, use a classic experimental paradigm. We did not, for example, randomly assign programs or even particular programmatic components to schools and teachers; to do so would have violated what we have learned from the last 20 years of research on school change—that school staffs must be involved in creating the programs for which they will be held accountable. However, it is neither necessary nor desirable to invite each and every school to rediscover the wheel. Therefore, what we did was to offer school staffs a framework for making their own decisions about how they might redesign their reading program. The framework consists of a set of six components derived from research-based knowledge about how to build an effective reading program. These components include classroom reading instruction, school reading programs, reading interventions, school-home-community relations, school change processes, and professional development. Each component is made available to a school via an Internet-based multimedia program offering research summaries, readings, video clips of effective practice, and learning activities to guide local action. Support for implementation of the framework is provided to schools through assessment tools and the data obtained with those tools, an external facilitator, an internal leadership team, schoolwide efforts, and study groups that focus on implementing effective practices.

Method

Participants

Five schools served as project schools and used the CIERA School Change Framework in 1999–2000. Three of these schools continued with the project in 2000–2001, and six additional schools joined the project at this time. All were high-poverty schools, with 70–95% of the students qualifying for subsidized lunch. Across schools, 2–68% of the students were non-native speakers of English, and 67–91% of the students were members of minority groups. The 11 project schools were from eight different school districts spread

across a rural area in the southeastern U.S.; an eastern city; two small towns in the Midwest; a large city in the Midwest; and a large city in the southwestern U.S. In order to become a project school, at least 75% of the teachers in a building had to agree to participate in the project. In all schools, two teachers per grade were randomly selected and invited to participate in the classroom observations, interviews, and completion of instructional logs. If a teacher declined, children from her classroom remained in the school-level analyses. Because the grade levels within buildings differed, children in 7 schools came from grades K–5, in 3 schools from K–6, and in 1 school from grades K–3. Within the designated classrooms, teachers were asked to divide their classes into thirds (high, average, and low) in terms of perceived reading ability; children were then randomly selected from each third. In the fall, 9 children were randomly selected as target students: 3 each from the high, middle, and low thirds of the classroom continuum. In the spring of 1999–2000, due to resource limitations, 2 high, 2 average, and 2 low children were randomly selected from each classroom for post-testing. In the spring of 2000–2001 as many as possible of the 9 children per class who remained at the school were tested: the average was 8 children per class.

Student Assessments

The children who were randomly selected for participation were assessed in the fall and spring on a number of literacy measures, which varied depending on grade level. Assessments included a standardized reading comprehension test (grades 1–6) as well as tests considering letter-name knowledge (K–1), rhyme (K–1), phonemic awareness (K–1), word dictation (K–1), concepts of print (K–1), fluency (words correct per minute; Deno, 1985) (1–6), and writing (responding to a common prompt) (1–6). See Table 1 for details.

Table 1. Assessments Used in Years 1 and 2

Assessment Tool	Description	Fall	Spring
Letter Names	Letter Name subtest from the Emergent Literacy Survey (Pikulski, 1996). Students identified both lower- and upper-case letters.	K, 1	K
Phonemic Awareness	Rhyming subtest from Emergent Literacy Survey (Pikulski, 1996). Students were given a word and asked to say another word that rhymed with it. Nonsense words were acceptable. Total of 8 points.	K	
	Classroom Segmentation and Blending Test (Taylor, 1991). Children were given six words to blend (e.g., "What is /c/ /a/ /t/?"). Then they were given six more words and asked to identify the first, middle, and last sound they heard in each word (e.g., "What sound do you hear first in sad? What sound do you hear in the middle of sad? What sound do you hear at the end of sad?").	1	K

Table 1. Assessments Used in Years 1 and 2

Assessment Tool	Description	Fall	Spring
Concepts of Print	Concepts of print subtest from Emergent Literacy Survey (Pikulski, 1996). Children were asked to identify a letter, word, and sentence, as well as demonstrate knowledge of tracking during reading. Total of 8 points.	K	K
Word-Level Dictation	Graded lists from Right Start Project (Colt, 1997). Students were asked to write 15 dictated words. If they could write at least seven words correctly from the first list, they went on to a second list of 15 words. (Administered in a group.)	1	K
Writing Prompt	Michigan Writing Assessment (administered in a group). Students were asked to write to a prompt (e.g., "Tell about a favorite place."). The same prompt was used in the fall and spring. A scoring rubric was used to score papers from high (4) to low (1).	2–6	1–6
Instructional Reading Level	Basic Reading Inventory (Johns, 1997). Students read graded passages until they reached frustration level. Instructional reading level was determined to be the highest level at which they could read with better than 90% accuracy in word recognition.		1
Fluency	Words correct per minute on passage from Johns (1997). In fall, students read for 1 minute from a passage that was one level below grade level. In spring they read from a passage that was at grade level.	2–6	1–6
Retelling	Retelling of Johns (1997) BRI passage read in winter.		
Houghton Mifflin Baseline Test	Houghton Mifflin Baseline Test (narrative only, administered in a group). After reading a three-page story, students answered five short answer questions and five multiple-choice questions. Possible score is 20 (0, 1, or 2 points for each short answer question, and 2 points for each multiple choice question correct). A score of 0–10 is considered low, 11–15 average, and 16–20 high.	2–6	1–6
Standardized Reading Comprehension Test	Gates-MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000). Only the passage comprehension subtest was administered. Students read short passages and answered multiple-choice questions. (Administered in a group.)	2–6	1–6

In the fall, kindergarten children were individually assessed (Pikulski, 1996) on letter-name knowledge (students were asked to give the names of the

upper- and lowercase letters); concepts of print (students responded to 8 items dealing with concepts related to words, letters, sentences, tracking, etc.); and rhyming ability (students responded to eight items, each giving a word that rhymed with a prompt). In the spring, kindergarten students were individually assessed on letter-name knowledge, concepts of print, and rhyme. They also completed an individually-administered, 12-item phonemic segmentation and blending test, in which they segmented words into phonemes and blended phonemes into words (Taylor, 1991), and a group-administered word dictation test in which they wrote 15 pre-primer and 15 primer words (Colt, 1997).

In the fall, children in grade 1 were individually assessed on letter-name knowledge, and phonemic segmentation and blending, and children were assessed in small groups on word dictation. In the spring all students were individually assessed on reading fluency (in which students read aloud for 1 minute to obtain a score for the number of words read correctly in 1 minute; Deno, 1985) based on a grade-level passage from the Basic Reading Inventory (BRI) (Johns, 1997). In a group setting students took the reading comprehension subtest of the Gates-MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) and responded to a writing prompt in which papers were scored according to a 4-point scoring rubric (Michigan Literacy Progress Profile, 1998).

In the fall, children in grades 2–6 were individually assessed on fluency (words correct per minute) based on their reading of a BRI passage (Johns, 1997) that was one grade level below their grade placement. As a group they took the comprehension subtest of the Gates-MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) and a writing prompt. In the spring, all children were assessed on fluency using a passage at grade level (Johns, 1997), on reading comprehension (Gates), and on writing (using the same prompt as was used in the fall).

Each response to the writing prompt was scored by one person from a team of trained scorers according to a rubric. Twenty-five percent of the writing samples at each grade level were scored by a second scorer, with 83% agreement between the 2 scorers.

Use of the School Change Framework

Teachers voted by secret ballot on whether to participate in the school change project. Schools were eligible to join the project if 75% of the teachers in the building voted to participate. Staff agreed to meet for a minimum of 1 hour a month as a large group to work on the school change effort, and 1 hour a week, on average, in smaller and more focused study groups. A school leadership team made up of teachers, the principal, and an external facilitator (who was to spend a minimum of 8 hours a week in the building) was responsible for guiding the staff through the school change activities. Large-group activities were to include discussion and action on the school-wide reading program, early reading interventions, and parent partnerships, as well as on issues related to school change and professional development. Reports were also expected from the study groups. Small-group activities were to include within-grade and across-grade study groups which focused

on particular aspects of classroom reading instruction and student work (e.g. comprehension instruction, phonemic awareness instruction). Groups were encouraged to review the research on the CIERA School Change website; to download, read, and discuss articles on research-based practices related to their focus area; to view and discuss video clips of effective practice on the site; and to share video clips of their own practice. Members of study groups also raised issues, solved problems, and developed action plans related to their focus area to make changes in their classroom reading instruction.

In addition to these components, schools agreed to several other practices and commitments in this multi-year project: cross-grade collaboration; the development of a plan to involve parents as partners in the delivery of their school reading program; and the commitment to continue with the project for at least 2 years, addressing, across that time span, all six major categories of the school change framework.

Use of data emanating from the project was also an important component. At the beginning of Year 1, facilitators received a summary of the Beating the Odds research (Taylor et al., 2000) on characteristics of effective schools and teachers, which they were asked to share with the teachers in their schools. Teachers also completed a checklist asking about various topics they felt they should cover during the year on characteristics of effective schools and teachers; these topics were covered on the Internet site. The purpose of both of these activities was to help schools set priorities for study groups and large-group meetings.

At the beginning of the second year, returning schools received a summary of the Beating the Odds research and a personalized school report that focused on their performance on school and classroom variables, as compared to the mean of other schools in the study. Schools new to the project received a generic version of the school report that included the cross-school means for school and classroom variables. The report included correlations identifying the school and classroom factors which are related to growth in students' reading and writing ability. Finally, the teachers also completed a questionnaire about their perceptions regarding the presence of various school and classroom characteristics, and their opinions about where their school should focus its reform efforts during the upcoming year. This questionnaire—like its predecessor, the previous year's checklist—was tied to topics covered on the Internet site, and was designed to help schools set priorities for their study groups and overall reform efforts.

Documenting Program Characteristics and Classroom Practices

Teachers were interviewed in the fall, winter, and spring; principals were interviewed in the fall and spring. The interview data were used primarily to document program features and participant beliefs. Each interview lasted about 30 minutes.

Teachers meeting in study groups were asked to complete a common study group meeting form after each session and develop an action plan. The external facilitators were asked to keep brief monthly notes summarizing

the activities pertaining to the school change project that had transpired at their school. They were also asked to write an end-of-year report. The data from the notes, action plans, and end-of-year report were used to document the change process at the school level.

On three occasions (fall, winter, spring), each teacher who agreed to be in the data collection sample was observed for an hour during reading instruction time, to document their classroom practices in the teaching of reading. All observations were scheduled. The observers were trained to use the CIERA Classroom Observation Scheme (Taylor et al., 2000), and were expected to demonstrate at least 80% agreement with a "standard" coding at each of the seven levels of the coding scheme (Taylor et al., 2000), prior to conducting classroom observations.

The observation system (influenced by the work of Scanlon & Gelzheiser, 1992; Greenwood et al., 1995; and Ysseldyke & Christenson, 1993–96) combined qualitative note-taking with a more quantitatively-oriented coding process. The observer took field notes for a 5-minute period, recording a narrative account of what was happening in the classroom, including, where possible and appropriate, what the teacher and children were saying. At the end of the note-taking period the observer recorded the proportion of children in the classroom who appeared to be on task (i.e., doing what they were supposed to be doing). They then coded the three or four most salient literacy events (Category 4 codes) that occurred during that 5-minute episode. For every Category 4 event, the observer also coded who was providing the instruction (Category 1), the grouping pattern in use for that event (Category 2), the major literacy activity (Category 3), the materials being used (Category 5), the teacher interaction styles observed (Category 6), and the expected responses of the students (Category 7). An example of a 5-minute observational segment is provided in Table 2. (See Table 3 for a list of the codes for all the categories.) In Table 2, the codes "c/s/r" refer to categories 1–3, and codes "r/t/a/r", "wr/t/c/or-I", and "v/t/r/or" each refer to categories 4–7.

Table 2. Sample of observational notes

9:38---Small group continues. T is taking running record of child's reading. Others reading familiar books. Next, T coaches boy on sounding out "discovered." Covers up word parts as he says remaining parts. T: "Does that make sense?" T: "What is another way to say this part ['cov' with short 'o']"? T passes out new book: My creature. T has students share what the word "creature" means. Ss: animals, monsters, dinosaurs, Dr. Frankenstein. 11/12 OT (On Task) C/s/r r/t/a/r wr/t/c/or-I (indv) v/t/r/or

Table 3. Codes for Classroom Observation

Level 1—Who	Code
Classroom teacher	c
Reading specialist	r
Special education	se
Other specialist	sp
Student teacher	st
Aide	a
Volunteer	v
No one	n
Other	0
Not applicable	9
Level 2—Grouping	**Code**
Whole class/large group	w
Small group	s
Pairs	p
Individual	i
Other	0
Not applicable	9
Level 3—General Focus	**Code**
Reading	r
Composition/writing	w
Spelling	s
Handwriting	h
Language	l
Other	0
Not applicable	9
Level 4—Specific Focus	**Code**
Reading connected text	r
Listening to text	l
Vocabulary	v
Meaning of text, lower m1 for talk m2 for writing	 m1 m2
Meaning of text, higher m3 for talk m4 for writing	 m3 m4
Comprehension skill	c
Comprehension strategy	cs
Writing	w
Exchanging ideas/oral production	e/o
Word ID	wi
Sight words	sw

Table 3. Codes for Classroom Observation

Phonics p1 = letter sound p2 = letter by letter p3 = onset/rime p4 = multisyllabic	 p1 p2 p3 p4
Word recognition strategies	wr
Phonemic awareness	pa
Letter ID	li
Spelling	s
Other	o
Not applicable	9
Level 5—Material	**Code**
Textbook, narrative	tn
Textbook, informational	ti
Narrative trade book	n
Informational trade book	i
Student writing	w
Board/chart	b
Worksheet	s
Oral presentation	o
Pictures	p
Video/film	v
Computer	c
Other/not applicable	o/9
Level 6—Teacher Interaction	**Code**
Tell/give info	t
Modeling	m
Recitation	r
Discussion	d
Coaching/scaffolding	c
Listening/watching	l
Reading aloud	ra
Check work	cw
Assessment	a
Other	o
Not applicable	9
Level 7—Expected Pupil Response	**Code**
Reading	r
Reading turn-taking	r-tt
Orally responding	or
Oral turn-taking	or-tt
Listening	l
Writing	w

Table 3. Codes for Classroom Observation

Manipulating	m
Other/not applicable	o/9
Number of students on task/ total number of students in room	

At the end of the observation, the observer wrote a summary addressing seven key features of the classroom ecology: (a) the general instructional approach used in the classroom, instructional sequences observed, approaches to word recognition, vocabulary, and comprehension instruction; (b) curriculum materials used; (c) teacher's style of interacting with the children; (d) teacher's grouping practices, and activities of children not with the teacher; (e) student engagement; (f) classroom management; and (g) classroom climate.

The observations were used as a source of data for individual teachers. In February of Year 1 teachers were invited by letter to receive, upon request, copies of their first two classroom observations and an explanation of the codes used in the observations. At one school, 75% of the teachers asked for copies of their observations; at three more schools, requests for feedback ranged from 42 to 50% of teachers; and at the fifth school, only one teacher out of 14 asked for feedback. In Year 2, based on requests from teachers for regular feedback related to their observations, teachers received a copy of each observation, a description of the codes, a brief summary of research related to the major coding categories being analyzed for the project (e.g., incidence of whole group instruction, and incidence of higher-level questioning; see page 17), and a table summarizing the codes from their observations (e.g., the incidence of whole-group instruction, the incidence of higher-level questioning, etc.) and comparing them to the means at their grade level across all schools. External facilitators received training in how to interpret observations so that they, in turn, could help teachers understand the information contained in these observations without making the interpretations for them. Teachers were encouraged to go to the facilitators with questions.

In addition to the observations, in 2000–2001 teachers also completed six logs, one each for a high-, average-, and low-ability student, for an entire week in the winter, and once again for the same three students in the spring. Teachers used the log to record time spent on various literacy activities, types of texts and materials used, and grouping practices.

Data Analyses

Data from the interviews were used to document the characteristics of various school-level factors at each school site, as well as the extent of the reform effort at that school. The meeting notes and action plans completed by the study groups, along with the monthly notes and end-of-year report completed by the facilitators, were used to further describe the extent of implementation of the change process at the project schools.

We applied a coding rubric to the interviews in order to evaluate teachers' perceptions of the degree to which the following factors, previously found to be important in effective schools (see pp. 1–3), existed in their schools: (a) building collaboration in the delivery of reading instruction; (b) links to parents; (c) reflection and change pertaining to instruction; (d) collaborative professional development; and (e) strong building leadership (and the extent to which this leadership was invested in the teachers, as well as the principal).[2]

Table 4. Summary Data from the Teacher Interviews and Descriptions of Categories Analyzed

TEACHER PERCEPTIONS	MEAN RATING FOR ALL SCHOOLS (BASED ON 4-POINT RUBRIC, IN WHICH 0 = LOW AND 3 = HIGH)
Links to Parents	1.89 (.11)
Collaboration	1.85 (.36)
Professional Development	1.84 (.32)
Reflection on Teaching	1.86 (.27)
Building Leadership	1.83 (.23)
Schoolwide Assessment System	1.99 (.03)
Total	11.3 (.9)

Each teacher's set of three interviews was used to rate school-level factors (collaborative leadership, building collaboration in the teaching of reading, reflection on teaching, collaborative professional development, and links to parents) on a 4-point scale, which was designed to capture the strength or degree to which each factor was perceived to be present in that school: (0 = very low perception, 1 = low, 2 = moderate, and 3 = high). This coding rubric is presented in Table 4. All of the interviews were coded by one member of the research team. A second team member independently coded the interviews from a random sample of 25% of the teachers; the mean agreement on overall rubric scores was 87% across the five variables.

The five ratings were summed to generate a school effectiveness score for each school in the study. The 11 schools from Year 2 and the three schools from Year 1 had a mean school effectiveness score of 8.3, (SD = 1.7), for a range from 5.4 to 10.2 (out of a total possible score of 15).

Although schools had agreed, in principle, to the conditions for the study, they exhibited considerable variability in their degree of adherence to the reform framework. Factors considered important to the reform included the following: (a) meeting for 1 hour a week in study groups; (b) meeting in cross-grade groups; (c) reflecting on teaching in study groups; (d) considering research-based "best practices" in study groups; (e) completing action plans in study groups; (f) selecting substantive topics for study; (g) maintaining topics over time; (h) meeting as a whole faculty once a month to discuss

---

2. The category of effective assessment practices was originally included in the rubric, but the questions in the interviews on assessment proved to be too few in number to provide us with useful information in this category, so it was dropped from the rubric.

reform efforts; (i) working on parent partnerships and making effective use of the external facilitator; and (j) making effective use of the internal leadership team. Using the comments of each teacher across the three interviews, the study group meeting notes, study group action plans, facilitator logs, and the end-of-year reports, we built a scale indicating the degree to which a school was perceived to be implementing the various components of the school change framework (see Table 5). We then calculated a mean reform effort score for each school. The mean score was 4.2 (SD = 1.6) out of a possible 10 points. One member of the research team rated each school on each of the 10 dimensions of implementing the reform. A second member of the research team also read through the artifacts and rated each school. The Pearson correlation coefficient across the two scorers' ratings was .92.

Table 5. Reform Implementation Rubric

1. Meeting 1 hour per week in study groups (at least 80% of the time).
2. Meeting in cross-grade study groups (at least 80% of the time).
3. Reflecting on instruction and student work (demonstrated at least 80% of the time).
4. Considering research-based practices (demonstrated at least 80% of the time).
5. Being guided by action plans (yes or no).
6. Sticking with substantive topics for 3–4 months or more (yes or no).
7. Meeting once a month as a whole faculty to share and set goals (at least 80% of the time).
8. Working on a plan to involve parents as partners (yes or no).
9. Effective use of an external facilitator (yes or no).
10. Effective use of an internal leadership team (yes or no).

Note: One point was awarded for each of the reform components if the criteria in parentheses for a particular component were judged to be met.

Coding the observations. As the first author of this paper visited research sites, she joined each observer in an observation, in order to establish inter-rater reliability data on the observation coding scheme. Across these 12 observations, agreement with the senior author ranged from 77% to 98%: 98% at Level 2 (grouping), 87% at Level 3 (major literacy focus), 80% at Level 4 (specific literacy activity), 90% at Level 5 (material), 80% at Level 6 (teacher response), and 77% at Level 7 (student response).

An expert observer who had done many classroom observations using this scheme and helped to refine it read through all of the observations to further assess the degree to which observers were using the codes in a similar manner. For example, although decision rules had been established in order to help an observer distinguish between similar codes, one observer may have coded a teacher's reference to the main idea of a story as a comprehension skill, while another observer might have coded a very similar exchange as a higher-level question about the story. The expert observer did not code the observations "blind." Instead, she recorded a different code only if she could not agree with the observer's code after reading the narrative description of a particular 5-minute segment. For a random sample of 10% of the observations, the agreements between the observers and expert observer at each of the levels of coding were measured as follows: 97% at Level 2 (grouping), 96% at Level 3 (major literacy focus), 80% at Level 4 (literacy activity), 86% at Level 5 (material), 77% at Level 6 (teacher response), and 83% at Level 7 (student response). Since there was variability between the

observers and the expert, especially at Levels 4, 6, and 7, a decision was made to use the expert's codes for those instances in which the observer and expert disagreed, in order to ensure maximum consistency across the many observers.

A second expert reviewer, a member of the research team, read through the same random sample of 10% of the observations. The agreement between the first and second expert at each of the levels of coding was very high: 98% at Level 2 (grouping), 97% at Level 3 (major literacy focus), 93% at Level 4 (literacy activity), 97% at Level 5 (material), 91% at Level 6 (teacher response), and 93% at Level 7 (student response).

Certain aspects of the data from classroom observations (i.e., those found to be important in previous research—c.f., pp. 2–3) were analyzed in order to investigate the relationship between various classroom instructional practices and students' reading and writing ability. The classroom practices which were analyzed included the following (see Table 6 for descriptions of the categories):

1. *Whole-class/large-group*—the percentage of segments in which whole-class or large-group was coded.

2. *Small-group*—the percentage of segments in which small-group was coded.

3. *Phonemic awareness skill instruction*—the percentage of segments in which phonemic awareness instruction was observed, divided by the number of segments in which the level 3 code was designated as reading.

4. *Phonics skill instruction*—the percentage of segments in which an aggregate of phonics skill instruction was observed, divided by the number of segments in which the Level 3 code was designated as "reading." The aggregate variable was formed by summing across the following practices: letter-sound instruction, letter-by-letter decoding instruction, instruction in decoding by onset and rime, instruction in decoding multisyllabic words.

5. *Coaching in word recognition strategies*—the percentage of segments in which coaching in word recognition strategies during reading was observed, divided by the number of segments in which the Level 3 code was designated as "reading."

6. *Comprehension skills*—the percentage of segments in which comprehension skill instruction was observed, divided by the number of Level 3 reading segments coded.

7. *Comprehension strategies*—the percentage of segments in which comprehension strategy instruction was observed, divided by the number of Level 3 reading segments coded.

8. *Lower-level questioning or writing about text*—the percent of segments in which lower-level talking or writing about text was observed, divided by the number of Level 3 reading segments coded. An aggregate variable was formed by summing the data from lower-level oral questions about text and lower-level written questions about text.

9. *Higher-level questioning or writing about text*—the percentage of segments in which higher-level talking or writing about text was observed, divided by the number of Level 3 reading segments coded. An aggregate variable was formed by summing the data from higher-level oral questions about text and higher-level written questions about text.[3]

10. *Teacher-directed stance towards teaching*—the percentage of responses in which teachers were coded as engaged in telling or recitation, out of the total number of the following responses coded: telling, recitation, modeling, coaching, and listening/giving feedback.

11. *Student-support stance towards teaching*—the percentage of responses in which teachers were coded as engaged in modeling, coaching, listening/giving feedback, out of the total number of the following responses coded: telling, recitation, modeling, coaching, and listening/giving feedback.

12. *Students actively responding*—the percentage of responses in which children were coded as engaged in reading, writing, or manipulating, out of the total number of Level 7 responses coded, including reading, writing, manipulating, reading-turn-taking, oral turn-taking, and listening.[4]

13. *Students passively responding*—the percentage of responses in which children were coded as engaged in reading turn-taking, oral turn-taking, or listening to the teacher out of total number of Level 7 responses coded, including reading, writing, manipulating, reading-turn-taking, oral turn-taking, listening.

For the 10% sample of observations described above, the second expert reviewer agreed with the first about the codes which made up the variables used in the data analyses: 100% whole-group, 99% small-group, 95% vocabulary instruction, 91% phonemic awareness instruction, 91% phonics instruction, 94% coaching in word-level strategies, 96% asking lower-level questions, 82% asking higher-level questions, 100% comprehension skill instruction, 88% comprehension strategies instruction, 94% teacher-directed stance, 92% student-support stance, 95% active responding, and 97% passive responding.

---

3. Because word skill work, comprehension skill/strategy work, and questioning or writing about text were almost always coded when the general focus of the lesson was reading at Category 3, a decision was made to consider the incidence of these three different types of reading activities out of the total number of Level 7 responses coded (including reading, writing, manipulating, reading turn-taking, oral turn-taking, and listening).

4. Because all Level 7 codes were frequently coded, and because multiple Level 7 codes were almost always coded during a 5-minute segment, a decision was made to consider the incidence of active (reading, writing, manipulation) and passive (reading turn-taking, oral turn-taking, and listening) events out of the total number of Level 7 codes recorded.

Table 6. Description of Classroom Observation Categories Used in Data Analysis

VARIABLE	DESCRIPTION
Whole Class or Large Group*	All of the children in the class (except for one or two individuals working with someone else), or a group of more than 10 children. If there are 10 or less in the room, code this as a small group.
Small Group*	Children are working in two or more groups. If there are more than 10 children in a group, call this whole group.
Phonemic Awareness Instruction	
Phonics Instruction**	Students are focusing on symbol/sound correspondences (p1) or letter-by-letter decoding (p2) or decoding by onset and rime or analogy (p3), but this is not tied to decoding of words while reading. If students are decoding multisyllabic words, code as p4. The total number of phonics activities out of the total number of times that reading was coded at Level 3 was calculated.
Word Recognition Strategies **	Students are focusing on use of one or more strategies to figure out words while reading, typically prompted by the teacher.
Lower Level Text Comprehension (talk or writing about text)**	Students are engaged in talk (m1) or writing (m2) about the meaning of text that is engaging them in lower-level thinking. The writing may be a journal entry about the text, or may be a fill-in-the blank worksheet that focuses on the text's meaning (rather than on a comprehension skill or vocabulary words). The total number of "low level text comprehension" activities at Level 4 out of the total number of times reading was coded at Level 3 was calculated.
Higher Level Text Comprehension (talk or writing about text)**	Students are involved in talk (m3) or writing (m4) about the meaning of text that is engaging them in higher-level thinking. This is talk or writing about text that is challenging to the children, and which is at either a high level of text interpretation or goes beyond the text: generalization, application, evaluation, aesthetic response. Needless to say, a child must go beyond a yes or no answer (e.g., in the case of an opinion or aesthetic response). The total number of "high level text comprehension" activities at Level 4 out of the total number of times reading (as the major focus) at Level 3 was coded.
Comprehension Skill Instruction**	Students are engaged in a comprehension activity (other than a comprehension strategy) which is at a lower level of thinking (e.g., traditional skill work such as identifying main idea, cause-effect, fact-opinion).
Comprehension Strategy Instruction**	Students are engaged in use of a comprehension strategy that will transfer to other reading and in which this notion of transfer *is* mentioned (e.g., reciprocal teaching or predicting. If predicting were done, but transfer was not mentioned, this would be coded as "c").
Vocabulary Instruction**	Students are engaged in discussing/working on word meaning(s).
Reading Text**	Students are coded as reading (not reading turn-taking) at Level 7.
Narrative Text*	The number of segments in which a narrative textbook (tn) or narrative trade book (n) was coded, out of the total number of coded segments.
Informational Text*	The number of segments during which an informational textbook (ti) or information trade book (i) was coded as being used, out of the total number of segments coded.
Teacher-Directed Stance***	The teacher is coded as telling or giving children information or engaging them in recitation. The total number of times in which telling or recitation was coded was divided by the total number of responses that were coded at Level 6.
Student-Support Stance***	The teacher is coded as coaching, modeling, or watching and giving feedback. The total number of times in which coaching, modeling, and watching/giving feedback were coded was divided by the total number of responses that were coded at Level 6.
Active Responding****	Children are engaged in one or more of the following Level 7 responses: reading, writing, oral responding, and manipulating. The total number of "active responding" codes out of the total number of Level 7 responding codes was calculated.

Table 6. Description of Classroom Observation Categories Used in Data Analysis

VARIABLE	DESCRIPTION
Passive Responding****	Children are engaged in one or more of the following Level 7 responses: reading turn-taking, oral responding turn-taking, listening. The total number of "passive responding" codes out of the total number of Level 7 responding codes was calculated.
Time on Task	At the end of the 5-minute note-taking segment, the observer took a count of the number of children in the room who appeared to be engaged in the assigned task out of all the children in the room. If a child was quiet but was staring out of the window or rolling a pencil on their desk, this was not counted as being on task.

*Percent of time (5-minute segments) coded.
**Percent of all reading segments coded.
***Percent of all codes for teacher interaction.
****Percent of all codes for student responding.

Statistical analysis. Hierarchical linear modeling (HLM; Raudenbush, Bryk, & Congdon, 2000) was used to investigate the impact of school-level and classroom-level characteristics on students' reading growth. Descriptive analyses were also conducted to elaborate on the quantitative findings.

HLM is a method of completing regression at multiple levels. The analyses in this study employed a two-level HLM model in which students were nested within classrooms or schools. Schools and classrooms were analyzed separately because there were different numbers of students at the school level than at the classroom level, since students whose teachers had declined to participate in the classroom observations were still included in the assessments for the school-level analysis. The number of schools was also insufficient to obtain stable results from a three-level model, in which students are nested within classrooms and classrooms are nested in schools.

HLM essentially estimates a regression within each school or classroom and combines these to see if they point to a common regression across schools or classrooms. When regressions (either the intercepts or slopes) vary across schools, then we can examine the school-level or classroom-level characteristics that may explain such variation. This is a common method for evaluating school-level and classroom-level factors and their effects on student outcomes. A simple regression would be inappropriate in these situations, since it would violate the independence assumption.

HLM also partitions variance components across levels, providing an estimate of variance in student performance within and between classrooms or schools. An unconditional HLM is one without an explanatory variable that allows us to answer the question: how much variance in student outcome can be attributed to systematic differences between classrooms and schools on specific factors? This analysis is equivalent to a random-effects analysis of variance. Estimations using HLM rest on assumptions similar to multivariate multiple regression.

Because of the improved estimation enabled by HLM, including the use of maximum likelihood and empirical Bayes estimates, interpretation of statistical results can be broadened to include a larger *p* value associated with statistical tests. Furthermore, statistical results with *p* values at or near 0.10 should be included in interpretation and explored in further studies with smaller numbers of cases (e.g., with fewer teachers or schools) because

such results indicate that there are relationships which merit further exploration. For a more complete description of estimation of HLM, see Bryk and Raudenbush (1992, pp. 32–56). HLM is recognized as a standard program for estimating multi-level models (Bryk & Raudenbush, 1992; Kreft & De Leeuw, 1998).

Results

Variations Across Schools in School-Level Characteristics

School effectiveness. Students' fall and spring scores are presented in Table 7. We began by running a two-level HLM analysis investigating the relationship between students' spring fluency scores and school effectiveness factors (See Table 8). This analysis was based on data from 877 students across all 11 schools (including the students from the three schools that were in the study during both Years 1 and 2). Our HLM analysis revealed that the school-level factor of collaborative leadership accounted for 29% of second- through sixth-graders' variance in spring fluency scores (number of words read correctly in one minute), after accounting for fall scores. This means that of the total variation in students' scores, 71% could be attributed to differences between students after accounting for fall scores irrespective of school, and 29% could be attributed to between-school differences. Differences in collaborative leadership scores in turn accounted for 24% of between-school variance.

Across all schools, the mean school fluency score was 104.5 words correct per minute. One way of gauging the influence of collaborative leadership is to note that for every additional point scored on the collaborative leadership scale, a school's mean fluency score showed an increase of 27.4 words correct per minute. If we note that students increased their scores by an average of 20 words correct per minute per year (see Table 7) and that school scores on the collaborative leadership scale ranged from 1.1 to 1.9 with a mean of 1.7 (out of a high score of 3), then we can surmise that, at least in principle, a school gaining one additional point on the collaborative leadership scale could make up a year's worth of fluency performance.

The HLM analysis of spring writing scores among second- through sixth-graders (see Table 9) revealed that the school-level factor of collaborative leadership accounted for 25% of the variance in students' spring writing after accounting for fall scores. Collaborative leadership scores accounted for 40% of the between-school variance.

Across all schools, the mean school writing score was 1.96 on a four-point scale. A way of gauging the importance of the 42% of variance contributed by collaborative leadership is to note that for every additional point on the collaborative leadership scale a school's mean writing score showed an increase of .85. After including fall scores, HLM analyses investigating the relationship between school effectiveness scores or subscores (e.g., collaborative leadership) and students' spring reading comprehension were not significant.

Table 7. Means and Standard Deviations for Student Scores K-6*

Assessment Tool	Grade	N	Fall	Spring
Letter Names	K	191	26.88	47.64
	1	199	(18.06)	(7.81)
			49.06	
			(6.02)	
Rhyme	K	191	3.09	5.39
	1	200	(3.22)	(3.00)
			5.23	
			(3.01)	
Phonemic Awareness	K	191		5.80
	1	200	6.59	(4.58)
			(4.59)	
Concepts of Print	K	191	4.25	6.84
	1	200	(2.58)	(1.64)
			7.29	
			(1.70)	
Word-Level Dictation	K	195		10.39
	1	213	14.72	(8.82)
			(9.58)	
Fluency	1	188		47.89
	2	200	52.22	(32.00)
	3	189	(31.74)	81.30
	4	192	76.51	(33.94)
	5	191	(30.69)	89.84
	6	68	102.73	(31.55)
			(36.59)	120.13
			123.23	(38.07)
			(42.50)	138.42
			130.44 (37.08)	(43.15)
				129.40 (44.86)
Gates Comprehension (NCE)	1	176		48.16
	2	214	40.05 (19.20)	(17.82)
	3	219	37.40 (14.03)	41.91
	4	190	35.68	(19.41)
	5	191	(17.68)	37.86
	6	68	36.11	(15.78)
			(18.83)	35.83
			41.98	(17.13)
			(18.76)	38.75
				(17.57)
				47.22
				(16.22)

Table 7. Means and Standard Deviations for Student Scores K-6*

Assessment Tool	Grade	N	Fall	Spring
Writing	1			
	2	211	1.61	1.89
	3	205	(.60)	(.72)
	4	154	1.76	1.92
	5	176	(1.55)	(.73)
	6	50	1.70	2.01
			(.67)	(.80)
			1.96	2.18
			(.70)	(.79)
			1.92	2.18
			(.72)	(.92)

*Only three of 11 schools had sixth-grade students. However, one of these schools had been with the project for 2 years, so 4 cohorts of grade 6 students were included in the data analyses.

Table 8. Grades 2–6 Reading Fluency with Collaborative Leadership

Initial Random Effects	Variance Component	% Variance Between		
School Means	226.11	29		
Fall Score Slope	.01			
Student Residual	560.07			
Total	786.19			
Final Random Effects		**% Variance Accounted for by Model**		
School Means	172.46	24		
Fall Score Slope	.009			
Student Residual	560.05			
Final Fixed Effects	**Coefficient**	***t*-ratio**	***df***	***p*-value**
Intercept (Grand Mean)	104.54	28.88	12	.000
Collaborative Leadership	27.39	2.22	12	.046
Fall Score	.82	25.31	13	.000

As reported earlier, all teachers in each building had completed a two-part self-study questionnaire during Year 2. The first part of the questionnaire dealt with teacher's perceptions of various building- and classroom-level factors within their school; the second part dealt with their opinions of the school's needs in regards to its change efforts. In the first part of the questionnaire, teachers rated each of 38 items on a scale from 1 (strongly disagree) to 5 (strongly agree). These items dealt with school change, climate, and leadership; professional development; schoolwide decisions about reading instruction; classroom reading instruction; reading interventions for struggling readers, and school-home-community connections.

Table 9. Grades 2-6 Writing with Collaborative Leadership

Initial random effects	Variance component	% Variance between		
School Means	.145	25		
Fall Score Slope	.029			
Student Residual	.416			
Total	.590			
Final random effects		**% Variance accounted for by model**		
School Means	.087	40		
Fall Score Slope	.030			
Student Residual	.416			
Final fixed effects	**Coefficient**	***t*-ratio**	***df***	***p*-value**
Intercept (Grand Mean)	1.96	23.67	12	.000
Collaborative Leadership	.85	3.02	12	.011
Fall Score	.35		13	.000

The school leadership ratings from our interviews were positively related to 12 of the 38 items from this first part of the questionnaire. The following is a list of the attributes which teachers saw as salient in cases where they perceived their schools to have strong leadership. The questionnaire items which were positively related to the interviews' school leadership ratings were:

- good trust and rapport;
- a culture of self-improvement at school;
- a good understanding of the process of school change;
- shared decision making;
- shared leadership;
- working together within a building on professional development;
- putting children first when reallocating resources;
- reducing negative elements in reading program;
- knowing where to get help to improve reading instruction;
- using student assessments to change instruction;
- having research-based reading interventions in place; and
- having reading interventions beyond grade 1.

Reform effort. We ran two-level HLM analyses investigating the relationship between students' spring comprehension, fluency, and writing scores and the school reform effort score for the total sample of 877 students in the 11 project schools (with data from three schools included separately across Years 1 and 2). HLM analyses investigating the relationship between the reform effort score and students' spring reading comprehension, fluency, or writing scores (after including the relevant fall scores) were not significant.

Using our coding of the 10 components of reform implementation, we were able to take a look at which schools were successfully implementing various factors. On the whole, we found that schools were having an easier time holding weekly study group meetings than they were holding monthly large-group meetings to share information across study groups and deal with schoolwide reform issues. Generally, schools were having an easier time meeting in grade-level groups than in cross-grade groups. Finally, schools were having an easier time reflecting on instruction and student work in study groups than they were focusing on research-based topics for periods of 3 months or longer. Most schools had not yet turned to the reform component of working with parents as partners (see Table 10).

Table 10. Reform Effort Ratings

Reform effort variable	Percent of schools demonstrating this reform variable
Meeting 1 hour per week in study groups	90%
Meeting in cross-grade study groups	10%
Reflecting on instruction and student work	90%
Considering research-based practices	20%
Being guided by action plans	30%
Sticking with substantive topics for 3–4 months or more	20%
Meeting once a month as a whole faculty to share, etc.	60%
Working on a plan to involve parents as partners	40%
Effective use of external facilitator	40%
Effective use of internal leadership team	80%

Descriptive Data from the Classroom Observations

In this section, we highlight results from our classroom observations as useful data in their own right, irrespective of school and independent of their relationship to student growth. In a sense, the data capture the nature of classroom instruction in schools like the ones in which we spent a year observing teachers and testing children. We report results for grades K, 1, 2–3, and 4–6. Descriptive data on classroom practices are presented in Table 11. They also provide us with an opportunity to compare what was going on in these schools with the practices we observed two years earlier in our study of high-poverty, high-performing schools (Taylor et al., 2000).

Table 11. Incidence of Classroom Factors by Grade

	KINDERGARTEN	GRADE 1	GRADES 2–3	GRADES 4–6
n =	27	28	54	58
Whole Group*	.60 (.34)	.47 (.28)	.51 (.33)	.61 (.33)
Small Group*	.38 (.29)	.46 (.29)	.43 (.30)	.37 (.33)
Phonemic Awareness**	.18 (.22)	.07 (.08)	.01 (.03)	—
Phonics Instruction**	.54 (.44)	.30 (.26)	.09 (.14)	.02 (.07)
Word Recognition Strategies**	.05 (.12)	.17 (.13)	.11 (.14)	.05 (.08)
Vocabulary**	.10 (.08)	.17 (.18)	.26 (.21)	.30 (.21)
Comprehension Skills**	.08 (.12)	.09 (.15)	.11 (.14)	.13 (.16)
Comprehension Strategies**	.01 (.02)	.06 (.13)	.05 (.09)	.10 (.17)
Meaning of Text—Lower Level**	.23 (.21)	.30 (.19)	.45 (.23)	.47 (.23)
Meaning of Text—Higher Level**	.04 (.07)	.06 (.11)	.14 (.23)	.21 (.20)
Informational Text*	.03 (.07)	.04 (.08)	.12 (.16)	.17 (.19)
Narrative Text*	.34 (.24)	.50 (.21)	.55 (.24)	.48 (.29)
Teacher Directed Stance***	.63 (.15)	.64 (.13)	.67 (.11)	.71 (.12)
Student Support Stance***	.37 (.15)	.36 (.12)	.33 (.11)	.29 (.12)
Active Responding****	.40 (.14)	.37 (.14)	.31 (.15)	.27 (.14)
Passive Responding****	.60 (.14)	.63 (.14)	.69 (.15)	.73 (.14)

*Percent of segments coded out of all segments coded.
**Percent of segments coded out of all reading segments.
***Percent of responses coded out of total number of Level 6 responses.
****Percent of responses coded out of total number of Level 7 responses.

Grouping practices. Across all grades, whole-group instruction was coded more often than was small-group instruction except for grade 1, in which small-group was coded as often as whole-group instruction. In contrast, in our earlier study of primary grade reading instruction in schools that were beating the odds (Taylor et al., 2000), we found that small-group, rather than whole-group, instruction across grades 1–3 was characteristic of the most effective schools.

Reading instruction practices. Not surprisingly, word-level activities during reading were observed more in grades K–1 than in 2–3 or 4–6. Phonics instruction was coded for about half of the reading segments in kindergarten, and a third of the reading segments in grade 1, but only about 5% of the segments in grades 2–6. Phonemic awareness instruction was coded for 18% of the segments in kindergarten and 7% of the reading segments in grade 1. Coaching in word recognition strategies during reading was coded for only 5% of the reading segments in kindergarten and grades 4–6, but for approximately 15% of the segments in grades 1–3. Comprehension skill instruction and comprehension strategy instruction were seldom observed. These findings are similar to those from our study of primary grade reading instruction in schools beating the odds (Taylor et al., 2000), in which we found that word-level activities were infrequently observed in grade 3, and that comprehension skill or strategy work was seldom observed in grades 1–3. The findings on word skill activities also suggest that teachers are focusing on phonics instruction in kindergarten and first grade, which is compatible with the recommendation of the National Reading Panel Report (2000) that

"phonics instruction taught early proved much more effective than phonics instruction introduced after first grade" (section 2, p. 85).

A relatively small amount of higher-level questioning or writing related to stories read was observed across all grades. Lower-level questioning was coded more often. A similarly low incidence of higher-level questioning was found in our earlier study (Taylor et al., 2000). However, in that study we did find that teachers in the most effective schools were more frequently observed asking higher-level questions than teachers in the moderately effective and least effective schools.

Materials. Informational text was seldom a part of the lessons we observed at any grade level. In contrast, narrative text was coded much more frequently.

Teacher and student actions. Telling and recitation were major interaction styles of teachers in all grades, with telling coded for 50–58% of the segments and recitation for 60–65% of the segments in grades K–6. In contrast, coaching was only observed for 12–21% of the segments in K–6. Overall, teachers were observed using a teacher-directed stance (i.e., telling and recitation) for approximately 60–70% of the teacher responses in kindergarten through grade 6. In contrast, a student support stance (coaching, modeling, listening/giving feedback) was coded for 30–40% of the teacher responses in grades K–6. In our earlier study (Taylor et al., 2000), we also found that telling was a common style of interaction, with the least accomplished teachers preferring to tell children information, and the most accomplished teachers preferring an interactive coaching style.

Across all grades, students in the present study were coded as more often engaged in passive responding than in active responding. Passive responding, which included reading turn-taking (e.g., round robin reading), oral turn-taking, and listening to the teacher, was coded for 60–70% of the student responses in grades K–6. In contrast, active responding (reading, writing, and manipulating) was coded for about 40% of the segments in kindergarten and grade 1, and for about 30% of the segments in grades 2–6.

Descriptive Data from Instructional Logs

In Year 2 of the study, teachers kept logs for two weeks (one in the winter and one in the spring), in order to document all literacy instruction in their classrooms. Students in full-day kindergarten classrooms averaged approximately 154 minutes per day of literacy instruction and activities. In grades 1–3 the average increased to 159–169 minutes per day. Grade 4 students averaged 142 minutes per day and grade 5 students 123 minutes per day. Generally, students spent about twice as much time in whole- or large-group instruction as they did in small-group instruction. Students across grades 1–5 spent about 25 minutes per day in independent reading.

Students in kindergarten and grade 1 averaged between 21 and 26 minutes per day of phonics or phonemic awareness instruction; this tapered off to an average of 16 minutes per day in grade 2, and 9 minutes per day in grade 3. In grades 1–5 students only averaged from 10 to 15 minutes per day of com-

prehension skill or strategy instruction. The means for the various literacy activities are shown in Table 12.

Table 12. Means From Teacher Logs of Minutes Per Day Spent on Various Literacy Activities Grades K–5

	Half-Day K	Full-Day K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
n =	5	9	17	16	17	18	15
Phonics/Phonemic Awareness	20.64 (10.10)	26.44 (13.29)	22.45 (14.82)	16.29 (8.99)	8.99 (6.79)	6.96 (8.52)	5.14 (7.54)
Other Word-Level Instruction	7.11 (1.82)	15.97 (15.92)	11.89 (9.40)	8.57 (4.94)	7.52 (5.71)	3.69 (3.92)	3.31 (3.98)
Vocabulary	8.99 (6.02)	7.84 (6.63)	11.97 (6.57)	10.06 (3.09)	13.54 (6.63)	10.90 (5.63)	9.57 (7.07)
Comprehension Instruction	4.92 (3.78)	7.58 (7.63)	10.11 (5.98)	12.28 (6.70)	11.15 (6.01)	11.25 (5.59)	15.58 (11.52)
Talking or Writing About Text	9.53 (5.13)	10.52 (9.54)	14.04 (7.96)	17.72 (9.35)	16.59 (5.68)	13.75 (8.38)	16.03 (10.90)
Teacher Reading Aloud	15.29 (9.89)	17.30 (8.07)	13.65 (6.88)	13.95 (8.42)	8.52 (4.34)	11.29 (9.15)	11.70 (8.34)
Teacher-Directed Reading	4.41 (3.99)	10.61 (5.56)	13.00 (7.22)	12.24 (6.64)	11.75 (10.27)	17.76 (8.09)	8.92 (6.94)
Independent Reading	12.92 (6.30)	16.20 (11.26)	21.28 (10.56)	23.33 (16.84)	25.83 (12.01)	26.54 (13.58)	22.18 (18.88)
Writing	7.89 (2.13)	20.76 (13.40)	21.06 (12.85)	27.32 (14.89)	28.52 (21.30)	18.14 (14.12)	15.83 (13.50)
Spelling	.32 (.38)	6.59 (9.80)	11.66 (9.91)	15.09 (9.06)	17.26 (9.02)	16.63 (11.62)	8.32 (7.16)
Other	8.84 (5.34)	13.89 (14.34)	14.09 (12.11)	11.81 (13.35)	9.37 (7.30)	10.55 (12.79)	6.62 (10.21)
Large Group	55.75 (12.48)	70.59 (26.97)	70.84 (38.24)	74.71 (35.06)	46.26 (23.59)	51.67 (18.98)	45.18 (21.44)
Small Group	25.04 (11.71)	31.50 (20.73)	34.97 (21.57)	26.10 (17.20)	35.69 (15.42)	28.84 (15.75)	26.61 (19.50)
Pair	.51 (1.13)	7.49 (6.17)	8.14 (9.11)	12.24 (11.25)	10.39 (10.62)	8.45 (9.59)	8.67 (8.01)
Individual	14.28 (13.28)	32.84 (21.44)	41.51 (26.74)	52.21 (20.17)	60.20 (31.22)	50.50 (22.67)	38.88 (30.44)
One-on-one	2.24 (4.79)	11.30 (11.71)	9.75 (10.30)	3.38 (4.77)	6.52 (9.67)	3.00 (5.66)	3.86 (8.35)
Total	97.87 (14.01)	153.70 (42.37)	165.22 (42.01)	168.64 (50.27)	159.06 (48.36)	142.46 (42.24)	123.20 (45.36)

Students' Reading Growth and Teacher Practices Across Schools

Eight of 11 schools were in their first year of the reform during Year 2 of the project. We had made a decision at the beginning of this project not to look at changes in instruction during a school year with only three observations per teacher. Therefore, before turning to an examination of changes in teaching in the three schools which were in their second year of the project, we

decided to take a look at the relationships between teacher practices during literacy instruction and students' reading and writing growth at various grade levels, irrespective of school, to see what could be learned about effective reading instruction practices. As was the case for Year 1 data from schools which started their participation in Year 2, these analyses provided useful Year 2 data as the schools began Year 3. HLM analyses were conducted on the relationships between teacher practices and each of the major outcome variables: reading fluency, Gates comprehension, and writing. Grade 1 was analyzed separately from grades 2–3 since different fall scores (e.g., word dictation in grade 1 versus words correct per minute, Gates comprehension, or writing score in grades 2–3) were used as explanatory variables in the analyses. Grades 2–3 and 4–6 were analyzed separately in order to look for patterns in teaching practices, which might affect differences in reading growth between the primary and intermediate grades. Kindergarten was analyzed separately because different pre-test and post-test variables were used.

Fluency. The HLM analysis (see Table 13) for grade 1 (188 students and 27 teachers) revealed that after accounting for fall scores, 41% of the variance in spring fluency scores was between teachers. Thirty-five percent of the between-teacher variance was accounted for by the variables of higher-level questioning and small-group instruction. Students' mean spring fluency score was 47.8. For every 10% increase in the coding of higher-level questioning, students' fluency scores increased by an average of 8.8 words correct per minute. For every 10% increase in the coding of small-group instruction, students' fluency score increased by an average of 2.1 words correct per minute.

Table 13. Grade 1 Reading Fluency With Higher-Level Questioning and Small Group

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	346.89	41		
Fall Score Slope				
Student Residual	501.84			
Total	848.73			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	224.70	35		
Fall Score Slope				
Student Residual	501.56			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	47.77	14.26	24	.000
High-Level Questioning	87.70	2.65	24	.014
Small Group	20.95	1.78	24	.088
Fall Score	2.09	8.19	184	.000

For grades 2–3, our HLM analysis of 341 students and 52 teachers (see Table 14) revealed that after accounting for fall scores, 45% of the variance in spring fluency scores was between teachers. Of this between-teacher variance, teacher-directed stance (negatively related) and phonics instruction (negatively related) together accounted for 12% of the variance. In grades 2–3, students' mean spring fluency score was 83.0. For every 10% increase in

the coding of a teacher-directed stance, students' fluency score decreased by a mean of 4.0 words correct per minute. For every 10% increase in the coding of phonics instruction, students' fluency score decreased by 3.0 words correct per minute on average.

For grades 4–6, the HLM analysis on data from 397 students and 56 teachers (see Table 15) revealed that after accounting for fall scores, 49% of the variance was between teachers. Coaching students in word recognition strategies during reading, having students engaged in active responding, and asking higher-level questions after reading accounted for 13% of the variance between teachers. Students' mean fluency score was 127.4 words correct per minute. For every 10% increase in the coding of coaching in word recognition strategies, students' fluency score increased by 8.9 words correct per minute on average. For every 10% increase in the coding of active responding, students' fluency score increase on average by 5.4 words correct per minute. For every 10% increase in higher-level questioning students' fluency score increased by 5.4 words correct per minute on average.

Table 14. Grade 2-3 Reading Fluency With Teacher-Directed Stance and Phonics

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	244.59	45		
Fall Score Slope	.009			
Student Residual	297.87			
Total	541.87			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	215.66	12		
Fall Score Slope	.01			
Student Residual	296.89			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	83.00	36.83	49	.000
Teacher Directed Stance	-39.94	-1.90	49	.063
Phonics	-29.50	-1.81	49	.077
Fall Score	.91	24.71	51	.000

Reading comprehension. The HLM analysis (see Table 16) for grade 1 (175 students and 25 teachers) revealed that after accounting for fall scores, 32% of the variance in spring comprehension scores was between teachers. Twenty-six percent of the between-teacher variance was accounted for by the variable of higher-level questioning. Students' mean spring comprehension NCE score was 48.6. For every 10% increase in the coding of higher-level questioning, a students' comprehension NCE score increased by an average of 4.5.

For grades 2–3, the HLM analysis on data from 380 students and 53 teachers (see Table 17) revealed that after accounting for fall scores, 29% of the variance in spring comprehension NCE scores was between teachers. A teacher-directed stance (negatively related) accounted for 13% of this between-teacher variance. In grades 2–3 students' mean spring comprehension NCE score was 39.9. For every 10% increase in the coding of a

teacher-directed stance, students' comprehension NCE score decreased by an average of 2.7.

Table 15. Grade 4-6 Reading Fluency With Coaching in Word Recognition and Active Responding

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	454.08	49		
Fall Score Slope				
Student Residual	479.21			
Total	933.29			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	393.47	13		
Fall Score Slope				
Student Residual	479.15			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	127.35	44.15	53	.000
Coaching in Word Recognition	89.41	2.38	53	.023
Active Responding	54.04	2.50	53	.016
Fall Score	.85	26.93	393	.000

Table 16. Reading Comprehension With Higher-Level Questioning

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	78.60	32		
Fall Score Slope				
Student Residual	163.79			
Total	242.39			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	58.50	26		
Fall Score Slope				
Student Residual	163.77			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	48.62	26.92	24	.000
Higher-Level Questioning	45.25	2.64	24	.015
Fall Score	1.27	8.96	173	.000

Table 17. Grade 2-3 Reading Comprehension With Teacher-Directed Stance

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	60.58	29		
Fall Score Slope				
Student Residual	147.82			
Total	208.40			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	52.96	13		
Fall Score Slope				
Student Residual	147.96			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	39.93	33.63	51	.000
Teacher-Directed Stance	-26.64	-2.44	51	.018
Fall Score	.76	16.63	377	.000

Table 18. Grade 4-6 Reading Comprehension With Higher-Level Questioning, Whole Group, and Teacher-Directed Stance

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	78.71	39		
Fall Score Slope	.006			
Student Residual	121.82			
Total	200.60			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	59.82	24		
Fall Score Slope	.062			
Student Residual	122.91			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	38.75	32.80	52	.000
Higher-Level Questioning	15.45	2.80	52	.008
Whole Group	8.45	2.32	52	.015
Teacher-Directed Stance	-16.24	-1.71	52	.092
Fall Score	.61	11.97	55	.000

For grades 4–6, the HLM analysis on data from 395 students and 56 teachers (see Table 18) revealed that after accounting for fall scores, 39% of the variance in spring comprehension scores was between teachers. Teaching with a teacher-directed stance (negatively related), asking higher-level questions after reading, and providing whole-class or large-group instruction together accounted for 24% of the variance between teachers. Students' mean comprehension NCE score was 38.8. For every 10% increase in the coding of a teacher-directed stance, students' comprehension score decreased by an average of 1.6. For every 10% increase in the coding of higher-level questions, students' comprehension NCE score increased by an average of 1.5.

For every 10% increase in the coding of whole- or large-group instruction, students' comprehension score increased by an average of 0.8.

Writing. The HLM analysis (see Table 19) for grade 1 (163 students and 24 teachers) revealed that 33% of the variance in spring writing scores was between teachers, after accounting for fall scores. Eighty-three percent of the between-teacher variance was accounted for by the variables of comprehension strategies instruction and student-support stance (negatively related). Students' mean spring writing score was 2.0. For every 10% increase in the coding of comprehension strategies instruction, students' writing score increased by an average of 0.4. For every 10% increase in the coding of a student-support stance, a student's writing score decreased by an average of 0.1.

For grades 2–3, the HLM analysis on data from 348 students and 50 teachers (see Table 20) revealed that after accounting for fall scores, 33% of the variance in spring writing scores was between teachers. Asking higher-level questions accounted for 15% of this between-teacher variance. In grades 2–3, students' mean spring writing score was 1.9. For every 10% increase in the coding of higher-level questioning, students' writing score increased by an average of 0.1.

Table 19. Grade 1 Writing With Comprehension Strategies and Student Support Stance

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	.170	33		
Fall Score Slope				
Student Residual	. 338			
Total	.508			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	.029	83		
Fall Score Slope				
Student Residual	.340			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	2.01	34.38	21	.000
Comprehension Strategies	3.92	6.29	21	.000
Student Support Stance	-1.34	-2.96	21	.008
Fall Score	.024	3.61	159	.001

Table 20. Grade 2-3 Writing With Higher-Level Questioning

Initial Random Effects	Variance Component	% Variance Between		
Classroom Means	.184	33		
Fall Score Slope	.070			
Student Residual	.301			
Total	.555			
Final Random Effects		**% Variance Accounted for by Model**		
Classroom Means	.157	15		
Fall Score Slope	.068			
Student Residual	.302			
Final Fixed Effects	**Coefficient**	***t*-ratio**	***df***	***p*-value**
Intercept (Grand Mean)	1.88	29.42	48	.000
Higher Level Questioning	1.01	2.72	48	.010
Fall Score	.36	4.82	49	.000

For grades 4-6, the HLM analysis on data from 345 students and 52 teachers (see Table 21) revealed that after accounting for fall scores, 44% of the variance in spring writing scores was between teachers. Coaching students in word recognition strategies during reading and asking higher-level questions after reading accounted for 12% of this between-teacher variance. Students' mean writing score was 2.1. For every 10% increase in the coding of coaching in word recognition strategies, students' writing score increased by an average of 0.2. For every 10% increase in higher-level questioning, students' writing score increased by an average of 0.1.

Table 21. Grade 4-6 Writing With Coaching in Word Recognition and Higher-Level Questioning

Initial Random Effects	Variance Component	% Variance Between		
Classroom Means	.278	44		
Fall Score Slope				
Student Residual	.353			
Total	.631			
Final Random Effects		**% Variance Accounted for by Model**		
Classroom Means	.246	12		
Fall Score Slope				
Student Residual	.354			
Final Fixed Effects	**Coefficient**	***t*-ratio**	***df***	***p*-value**
Intercept (Grand Mean)	2.06	26.90	49	.000
Coaching in Word Recognition	1.74	1.83	49	.073
Higher-Level Questioning	.77	2.01	49	.050
Fall Score	.78	4.63	341	.000

Kindergarten. For phonemic segmentation and blending, the HLM analysis on data from 182 students and 26 teachers (see Table 22) revealed that after accounting for fall letter-name scores, 36% of the variance was between teachers. Comprehension strategies instruction, phonemic awareness instruction, and phonics instruction (negatively related) accounted for 49%

of the between-teacher variance. Students' mean phonemic awareness score in spring was 5.7. For every 10% increase in the coding of comprehension strategies instruction, students' phonemic awareness score increased by an average of 4.9. For every 10% increase in the coding of phonemic awareness instruction, students' phonemic awareness scores increased by an average of 0.6. For every 10% increase in the coding of phonics instruction, students' phonemic awareness score decreased by an average of 0.2.

For concepts of print, the HLM analysis (see Table 23) revealed that after accounting for fall concepts of print scores, 24% of the variance was between teachers. Small-group instruction and phonemic awareness instruction accounted for 47% of this between-teacher variance. Students' mean spring concepts of print score was 6.8. For every 10% increase in the coding of small-group instruction, students' concepts of print score also increased by an average of 0.1. For every 10% increase in the coding of phonemic awareness instruction, students' concepts of print score increased by an average of 0.1.

Table 22. Kindergarten Segmentation and Blending With Comprehension Strategies, Phonemic Awareness, and Phonics

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	6.44	36		
Fall Score Slope				
Student Residual	11.21			
Total	17.65			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	3.26	49		
Fall Score Slope				
Student Residual	11.15			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	5.67	12.94	22	.000
Comprehension Strategies	49.54	2.60	22	.017
Phonemic Awareness	6.09	2.90	22	.009
Phonics	-2.01	-1.89	22	.072
Fall Score	.12	7.39	177	.000

Table 23. Kindergarten Concepts of Print With Small-Group and Phonemic Awareness

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	.565	24		
Fall Score Slope	.050			
Student Residual	1.69			
Total	2.31			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	.300	47		
Fall Score Slope	.041			
Student Residual	1.73			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	6.78	46.25	23	.000
Small Group	.88	1.99	23	.058
Phonemic Awareness	1.44	2.36	23	.027
Fall Score	.287	4.56	25	.000

For word dictation, the HLM analysis (see Table 24) revealed that after accounting for fall letter-name scores, 44% of the variance was between teachers. Phonemic awareness instruction and comprehension strategies instruction accounted for 31% of the between-teacher variance. Students' mean spring word dictation score was 10.4. For every 10% increase in the coding of phonemic awareness instruction, students' word dictation score increased by an average of 1.1, and for every 10% increase in the coding of comprehension strategies instruction, students' word dictation score also increased by an average of 0.8.

For rhyme, the HLM analysis on data from 182 students and 26 teachers (see Table 25) revealed that 21% of the variance was between teachers after accounting for fall rhyme scores. Phonemic awareness instruction and phonics instruction (negatively related) together accounted for 56% of the between-teacher variance. Students' mean rhyme score in spring was 5.4. For every 10% increase in the coding of phonemic awareness instruction, students' rhyme score increased by an average of 0.4. For every 10% increase in the coding of phonics instruction, students' phonemic awareness score decreased by an average of 0.2.

Table 24. Kindergarten Word Dictation With Phonemic Awareness and Comprehension Strategies

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	27.32	44		
Fall Score Slope				
Student Residual	34.40			
Total	61.72			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	18.87	31		
Fall Score Slope				
Student Residual	34.40			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	10.36	10.77	23	.000
Phonemic Awareness	11.21	2.50	23	.020
Comprehension Strategies	80.70	1.94	23	.064
Fall Score	.261	9.77	182	.000

Table 25. Kindergarten Rhyme With Phonemic Awareness and Phonics Instruction

INITIAL RANDOM EFFECTS	VARIANCE COMPONENT	% VARIANCE BETWEEN		
Classroom Means	1.68	21		
Fall Score Slope				
Student Residual	6.29			
Total	7.97			
FINAL RANDOM EFFECTS		% VARIANCE ACCOUNTED FOR BY MODEL		
Classroom Means	.735	56		
Fall Score Slope				
Student Residual	6.30			
FINAL FIXED EFFECTS	COEFFICIENT	*t*-RATIO	*df*	*p*-VALUE
Intercept (Grand Mean)	5.42	21.29	23	.000
Phonemic Awareness	3.79	3.09	23	.006
Phonics	-2.01	-3.22	23	.004
Fall Score	.359	5.55	178	.000

For spring letter-name knowledge, the HLM analysis on data from 184 students and 26 teachers revealed that after accounting for fall scores, teachers accounted for 38% of the variance in students' spring scores. However, no classroom observations variables contributed significantly to this between-classroom variance.

Summary of 2-level HLM classroom results. Across grades 1–6 the HLM analyses revealed that the asking of higher-level questions was positively related to students' reading and/or writing growth. A relatively high level of phonics instruction was not found to be helpful for students' growth in fluency in grades 2–3, or to their phonemic awareness development in kindergarten; but phonemic awareness instruction was found to be related to

students' spring emergent literacy scores in kindergarten. Coaching in word recognition strategies during reading was related to students' growth in reading fluency in grades 4–6. Comprehension strategy instruction was related to spring writing scores in grade 1, and to emergent literacy scores in kindergarten. A highly teacher-directed stance towards instruction was not found to be beneficial to students' reading growth in grades 2–6, whereas active responding was found to be beneficial to growth in reading fluency in grades 4–6. A high-level student support stance (e.g. modeling, coaching, watching/giving feedback) was not found to be beneficial to students' writing growth in grade 1. Small-group instruction was found to be beneficial in kindergarten and grade 1, whereas whole- or large-group instruction was found to be beneficial in grades 4–6.

Descriptions of More Helpful and Less Helpful Classroom Factors

To better explain the findings related to classroom factors, we provide descriptions of teachers who aptly illustrate the practices identified by the quantitative analyses as positive. We also provide examples from classrooms in which a heavily teacher-directed stance was apparent, in order to better describe a practice which was identified as less helpful to students' literacy growth. Below, we provide illustrations from the field notes, with direct quotes from teachers and students in italics.

Grade 1. The HLM analyses found that children grew more in comprehension and fluency when their teachers were coded as asking more higher-level questions than other teachers. Teachers who were more often observed teaching their students in small groups in first grade also had students who showed larger gains in fluency during the year. The teaching of comprehension strategies was found to be related to greater growth in writing in first grade children.

Ms. Hernandez. Ms. Hernandez[5], a first-grade teacher in the study, exemplifies many of these relationships in her classroom. She uses small-group instruction extensively. While she is with one group, the other students are in centers, where activities include writing words and word families, math, computer, library, and reading the room.

Ms. Hernandez focuses heavily on comprehension strategies and higher-level thinking as she teacher her first-graders. On one day, her students are reading a story. She introduces a GO chart with columns that are labeled "prediction," "vocabulary," "understanding," "interpretation," "connections," and "retelling," and prompts students to complete information on the chart. Prediction—"I think the story is about…." Students share various predictions based on the title and pictures. She asks them to check in their books to look for challenging vocabulary that they think they should add to the chart. Ms. Hernandez asks for a word that has the same meaning as "house." They add "cottage" to the chart. "Other interesting words? Another word for woods?" The students suggest "forest."

---

5. All names are pseudonyms.

Ms. Hernandez refers to the "Understanding—I noticed" column of the chart. A student suggests, "The giant does interesting things."

Ms. Hernandez refers to the "Interpretation—I wonder" column. She encourages the students to think about what happens next, to go beyond the story, to imagine what the characters could do together. "Think if you are one of the characters in the story how would you solve the problem? What connections can you make, what is the main thing you learned in the story?" A student explains that it's about friends.... Ms. Hernandez asks, "What maps can you use to help you retell the story?" Students suggest various graphic organizers, including story webs, circle maps, and a tree map. Ms. Hernandez asks how and why the class could use each one of these devices. One student suggests that they could use a bubble map to describe the character.

On another day, Ms. Hernandez is working with a small group studying informational texts. She refers to the GO Chart. She asks them to quickly review the steps without looking at the chart. Students respond "I think this story is about (prediction)," "I noticed important words in the story (vocabulary)," "I noticed (understanding)," "I wonder (interpreting)," "This reminds me (connections)," and "Maps/story/illustration (retelling)."

Mrs. Gleason. Mrs. Gleason, also a first-grade teacher, uses whole-group instruction about one third of the time and small-group activities about two thirds of the time. Like Ms. Hernandez, Mrs. Gleason often engages her first-grade students in higher-level thinking. On one day she has her students working on dog reports. They have collected information from multiple sources and are putting similar notes together. Writing the report requires higher-level thinking, as the children write down ideas that go together and follow an organizational pattern.

On a different day, after the students have read *George shrinks*, the small group is making a list of things they do at school—math, drawing, journal writing, eating lunch, special activities, feeding the bunny. Mrs. Gleason asks the students to think about how they would do these things if they were small. Kou says, "We could break the pencil and just write with the point." Mrs. Gleason asks, "How could we drink milk?" Kou replies, "Get a ladder?" Mrs. Gleason has students write, "If I were small, I'd need a toolkit to help me." Then she asks them to add sentences telling what things they would have in their toolkit, and how they would use these things. While the children are writing, Mrs. Gleason circulates through the class, helping individual students with their ideas and coaching them with spelling words.

Ms. Metcalf. Ms. Metcalf primarily teaches through small group instruction, but unlike Ms. Hernandez and Mrs. Gleason, she typically does a lot of the work for her children, illustrating a highly teacher-directed stance towards instruction. When she reviews the "magic *e*" rule with a small group, she says, "Today we'll review the silent *e*." She writes "fin," adds "e" and tells the children the rule, "*e* is silent and makes the vowel say it's name." She misses the opportunity to have the children tell her the rule. When she introduces new vocabulary words before reading a story, one of the words is "donkey." She tells them the word, and instead of asking them what a donkey is, she tells them the meaning, "donkeys have bigger ears than horses and are stubborn."

One day, Ms. Metcalf is reading a story to the children. As they discuss the story, she interjects her own ideas and summarizes for the group. "Why do you think they wrapped the bones?" A student answers, "So they wouldn't break." Ms. Metcalf confirms this correct response, offering a comparison with the need to wrap fragile items when packing. Ms. Metcalf also stops to repeat important points and clarify meanings, but she misses another opportunity to have the children do the talking. "So when you see a dinosaur skeleton in a museum, it probably has a few bones that are not originals, since some might have been missing, or broken. Fiberglass or plastic is used now."

Grades 2–3. In grades 2 and 3 the HLM analyses revealed that students had higher growth in reading comprehension when teachers were less often observed teaching with a highly teacher-directed stance (e.g., telling, recitation). Students showed greater fluency growth when their teachers were more often observed providing small-group instruction, and less often observed teaching phonics. Children showed greater growth in writing when their teachers were more often observed asking higher-level questions after reading.

Mrs. Schneiter—grade 2. Mrs. Schneiter, who teaches second grade, focuses on higher-level questioning. She models word recognition strategies frequently as students are reading. She primarily teaches through small-group instruction, and she also has those students who are not with a teacher work with a partner or small group to foster active involvement at their seats.

On one day, Mrs. Schneiter is helping students review what they should do when they are reading with a partner and the partner gets stuck on a word. She also writes on the board the things that students should do if they finish their work before she is finished with her group: 1) re-read the story, 2) write down some of Teeka's emotions and characteristics, and 3) write your favorite part—all higher-level thinking activities.

Later, she discusses characters' emotions with the group. "Now, I asked you to think about emotions and characteristics as you read the story." She makes two columns on the board. "So how do you think Teeka felt during the story?" The students give the following responses:

EMOTIONS	CHARACTERISTICS
sad, angry, mad	nice, kind, responsible
happy, frustrated	mean

She asks the children to take out their journals. Students are to write two sentences in their journals. They are to write about one emotion and one characteristic. Students are encouraged to look back at the story if they need to.

On a second day, Mrs. Schneiter talks to all groups. "I want you to take out your reading journals." She and an assistant move around the room as the children begin to write on the suggested topic: "What would you tell Arthur so that he could remember his speech? How would you solve the problem?"

Mrs. Stone—grade 3. In contrast, Mrs. Stone, a third grade teacher, focuses more on lower-level comprehension. Her style is highly teacher-directed, and she makes frequent use of recitation. One day, prior to working with a

group, she explains the seatwork activities. One group is to go to a listening center. Another is to think of five words that rhyme with "protect" and look up their definitions in the dictionary. This group is also to complete a worksheet in which they circle rhyming words—all low-level phonics activities.

In the reading group, Mrs. Stone asks a student to read aloud from *Frog's journal*, while other students follow along (reading turn-taking, which is coded as passive responding). When a student gets stuck on a word she tells it to them. She then asks low-level questions about a character in the section: "What will Frog want to hear? What did he need or want to tell Frog? Why was he still? What did he hear?"

On another day, Mrs. Stone is working with a reading group. "What's the title of the story?" One student reads "Think Positive." The teacher asks, "What does that mean?" A student responds correctly, but the teacher goes on to explain, "Positive means happy things, positive is the opposite of negative." She gives an example: in baseball, a batter needs to think positively. She then says that she will read the introduction, and begins to read while the students follow along. Overall this is a very teacher-directed lesson.

Grades 4–6. Asking students higher level questions after reading was related to their growth in fluency, comprehension, and writing. Coaching in word recognition strategies during reading was related to students' growth in fluency and writing. Having students engaged in active responding was also related to their growth in fluency. Teaching with a highly teacher-directed stance (negatively related) and providing whole-class or large-group instruction were additional classroom characteristics related to students' growth in reading comprehension.

Mr. May—grade 5. Mr. May, a fifth-grade teacher, provides many good examples of the "best practices" identified in the HLM analyses for grades 4–6. He provides more whole-group than small-group instruction, but students also work independently or in small groups for a fair amount of time in between whole-class segments of a reading lesson.

In a typical lesson, Mr. May begins by listing objectives for the reading hour on the board. Then he reviews the vocabulary words that students should be looking for as they read the story. Each word is introduced in the context of a sentence from the book; for example, "The Herdmans had music *blaring* in the background."

Mr. May stresses higher-level questions and active pupil involvement. As students prepare to read the next chapter in the *Best Christmas pageant ever*, he challenges them to think about what is happening next in the story. "Do you think the children should be in the play?" He takes a vote, which fosters active pupil involvement, and he has students defend their opinions. In their response journal, students are to answer the following questions: "Do you think the Herdmans will do a good job? Why or why not? How do you think the audience will react? Give evidence from the story."

Mr. May provides small-group instruction to struggling readers while the rest of the class is reading independently. He coaches the small group in word recognition as they read the chapter, and helps prepare them for the questions he'll be asking in the whole-group setting.

After Mr. May returns to the whole class and they have a discussion of what they wrote in their journals, he has students break into small groups. They are to work together in these groups to write the meanings for two vocabulary words and answer an assigned question. They put the answers on an overhead sheet in order to share their work with the rest of the class.

Mr. Burns—grade 4. In contrast Mr. Burns, a fourth-grade teacher, asks mostly lower-level questions, and his students are engaged in a considerable amount of passive responding. During one lesson, students are in three groups: one with an aide, one doing worksheets at their seats, and one with Mr. Burns. The teacher has them doing round-robin reading. When a student gets stuck on a word, Mr. Burns tells him the word. Questioning is a rapid fire of low-level questions: "Why did Mom miss Merritt? Did she use her cane when she had her guide dog with her? What did her daughter want her to do when she went on errands? What is the name of the school she is going to?"

On the second day, Mr. Burns works with the whole group. The introduction to the story takes 30 minutes. Prior to reading, Mr. Burns asks students about the meaning of words which they will come across when they read: *building, marketplace, celebration, foolishness, snickered, spice, vow, chessboard.* When the students don't know a word, he tells them what it means. Almost all of the talking is done by the teacher. Half of the students seem to not be engaged or listening. They don't look at the board or raise their hands. However, they are all well-behaved.

Kindergarten. Comprehension strategies instruction, phonemic awareness instruction, and phonics instruction (negatively related) significantly contributed to students' growth on a variety of emergent literacy measures. Small-group instruction contributed to growth in concepts of print.

Ms. Jackson. Ms. Jackson demonstrates all the characteristics of an effective kindergarten teacher, as determined from the HLM analyses. She teaches reading in small groups of four. She uses many word-level activities: the children make words with plastic letters, write the sounds that they hear in common words as they create simple books or write in their journals, and look for patterns in words like "neck" and "deck."

Instead of telling the children information, Ms. Jackson gets students actively involved in their lessons. When introducing a new book to a small group of kindergartners, Ms. Jackson says, "This story is going to take place somewhere different. Where do you think this one takes place?" Students look at the cover and point out a giraffe and a snake. One student says, "The zoo!" The children discuss the animals. Ms. Jackson asks, "Do these animals live on a farm?" The children respond. The teacher asks, "Have you ever seen a real lion? What do they eat?" One child says, "Animal food." Another child says "People." Ms. Jackson responds, "Do you think they eat people?" Several children share their opinions. The teacher asks, "What is the difference between a lion and a tiger?" Students discuss the fact that a lion has hair around its face. Following their discussion, Ms. Jackson has every child in the small group read the book aloud while she listens and coaches in word recognition strategies. For example, when a child is stuck on a word, Ms. Jackson says, "What does it start with? What could it be? Look at the picture." The child says "Cow." Ms. Jackson asks, "Is there more than one?" The child responds, "Two." Ms. Jackson asks, "How would you read this word then?"

Ms. Lawson. Ms. Lawson, another kindergarten teacher, also uses small groups for reading instruction, but she relies on lower-level questioning about text more than Ms. Jackson does. For example, after reading *The very hungry caterpillar* by Eric Carle, she says, "I'm going to ask five questions on the story I just read to you. What is the name of the story? Who was the main character in the story? What did he do in the story?" These questions are presented in rapid succession and the teacher quickly moves on after receiving the correct answers.

Analyses of Students' Reading/Writing Performance and Classroom Practices in Schools Completing Two Years in the Project

Three schools had each been in the project for 2 years. Separate analyses were conducted on student scores and teacher practices at these schools across the 2 years, in order to investigate the impact of the reform effort.

An analysis of covariance on students' spring comprehension scores across grades 2–6, with fall scores used as the covariate, revealed a significant effect for the school: $F(2,336) = 10.18$, $p < .001$; a significant effect for year in study (Year 1 or Year 2): $F(1,336) = 4.18$, $p = .04$; and a significant school by year-in-study interaction: $F(2,336) = 10.30$, $p < .001$. Bonferroni pairwise comparisons revealed that School 1 and School 2 had significantly higher adjusted spring comprehension scores than School 3. Overall, schools had higher adjusted spring comprehension scores in Year 1 than in Year 2.

However, School 1 had significantly higher adjusted spring comprehension scores in Year 2 than in Year 1: $F(1,97) = 10.44$, $p = .02$. There was no difference between Year 1 and Year 2 scores in Schools 2 and 3.

An analysis of the covariance on students' spring fluency scores across grades 2–6, using fall fluency scores as the covariate, revealed a significant effect for school: $F(2, 344) = 4.68$, $p = .01$. School 1 had significantly higher adjusted spring scores than School 3. There was no difference between Year 1 and Year 2 scores in any of these schools.

An analysis of the covariance on students' spring writing scores across grades 2–6, using fall writing scores as the covariate, revealed a significant effect for school: $F(2, 293) = 5.44$, $p = .005$; and significant school by year-in-study interaction: $F(2, 293) = 3.90$, $p = .02$. School 1 had significantly higher adjusted spring writing scores than Schools 2 and 3. School 3 had significantly lower spring writing scores in Year 2 than in Year 1. There was no difference between Year 1 and Year 2 scores in Schools 1 and 2.

An analysis was done of teaching practices across Years 1 and 2 at School 1, which had higher spring comprehension and fluency scores than School 3 after adjusting for fall scores, and higher spring writing scores than Schools 2 or 3 after adjusting for fall scores. School 1 also showed increased growth in spring reading comprehension from Year 1 to Year 2 in an analysis of covariance. Analyses revealed at least a 10% difference between Years 1 and 2 in

teacher observations in grades 2–6 for the following factors: decrease in whole-group instruction, increase in small-group instruction, increase in asking of higher-level questions, increase in comprehension strategies instruction, increase in teacher-directed stance, decrease in student support stance. Except for the increase in teacher-directed stance and decrease in student-support stance categories, all differences in classroom practices were in the direction which would be expected based on the research on effective classroom reading instruction, which had been shared with the teachers before Year 1 and between Years 1 and 2. These findings suggest that teachers at School 1 were making shifts in their classroom reading instruction and were positively influenced by the research on best practices in reading. The reform effort rating had stayed at 4 out of a possible 10 for Years 1 and 2. The school effectiveness rating increased from 6.8 to 9.4 between Year 1 and Year 2. Teachers had more positive perceptions about building collaboration, instructional reflection, professional development, and parent partnerships in Year 2 than they had had in Year 1.

The results were more mixed at School 2, with respect to changes in classroom practice. Analyses revealed that the following observations changed by at least 10% from Year 2 to Year 1: increase in whole-group instruction, decrease in small-group instruction, increase in coaching in word recognition strategies during reading, decrease in asking of lower-level questions, increase in asking of higher-level questions, increase in comprehension skill instruction, decrease in active pupil responding, increase in passive pupil responding. Some of these changes, such as increase in coaching in word recognition strategies and increase in asking of higher-level questions (and the corresponding decrease in asking of lower-level questions) moved in the direction that research would suggest was beneficial. Other changes moved in a direction that research would suggest was not beneficial. Again, we shared research on effective reading instruction with teachers from School 2 both before Year 1 and between Years 1 and 2.

The strongest positive change in classroom practices at School 2 was related to comprehension. Teachers had decided that improving students' reading comprehension should be one area of focus for Year 2. During Year 2, teachers did appear to shift to asking students more higher-level questions about what they had read than they had in Year 1. However, observations also indicated an increase in comprehension skill instruction rather than comprehension strategy instruction, although the latter has been found to be the most effective approach for increasing reading achievement (NRP, 2000). The reform effort rating had dropped from 4 to 3 between Years 1 and 2. There had been a change in principals between Years 1 and 2, and as a result teachers were meeting in study groups less often in Year 2 than they had in Year 1. However, on a positive note, the school effectiveness rating had increased from 8.4 to 9.8 during the same time, with this increase primarily due to teachers' perceptions of building collaboration and parent partnerships growing further in the positive direction in Year 2 than they had in Year 1.

At School 3, analyses revealed a decrease of more than 10% in the asking of higher-level questions from Year 1 to Year 2. No other classroom practices increased or decreased by 10% or more over this same period. As we had done at Schools 1 and 2, we shared the research on best practices with teachers at School 3 before Year 1, and again between Years 1 and 2. As was the case at School 2, this research did not make much of an impact on the

teaching of reading at School 3. However, School 3's reform rating did increase from 1 to 4 between Years 1 and 2, most likely because the new external facilitator was able to get study groups meeting more consistently in Year 2 than they had done in Year 1. The school effectiveness rating increased from 6.6 to 8.8 between Year 1 and Year 2, due to the efforts of the facilitator, a newly-hired reading specialist, and the principal. In Year 2 teachers also showed more positive perceptions about building collaboration, professional development, leadership, and parent partnerships.

Summary of changes in classroom instruction at schools that participated in the project for two years. Based on the evidence from just these three schools, we have to conclude that the effect of the reform effort on classroom reading instruction was mixed at the end of Year 2. It does appear that the reform effort is helping teachers at one school make changes in their classroom reading instruction according to research-based practices; increased growth in students' reading achievement was observed from Year 1 to Year 2. At the second school, which had established the improvement of students' reading comprehension as a schoolwide goal, we observed a shift towards higher-level questioning. In general, however, we did not see a shift toward more effective reading instruction practices at this school. The third school showed little evidence of a shift towards more effective reading instruction practices across Years 1 and 2.

Discussion

School-Level Findings

We often hear that effective school reform in reading—reform that significantly raises student achievement—takes dedication, hard work, and time. The results of our study confirm this assertion. Even though the vast majority of teachers at all schools in the project voted to engage in schoolwide reading reform, the schools clearly have a long way to go to raise their reading scores to "breaking the mold" levels, with current mean standardized reading comprehension scores across schools, for example, standing at 40 in grades 2 and 3, and at 37 in grades 4 and 5. If we consider the factors of building collaboration, professional development, instructional reflection and change, collaborative leadership, and parent partnerships, then schools had a mean effective school rating of 8.3 out of a possible score of 15, leaving considerable room for additional growth. However, the schools that joined the project *are* aspiring to beat the odds. Each of the three schools that have been with the project for 2 years increased their school effectiveness rating from Year 1 to Year 2. Our shared hope, as we continue to follow these schools over the next year, is that we will see significant changes toward more effective school-level and classroom-level practices and more success in reform efforts—changes which will be accompanied by increases in student performance on a range of reading measures.

Even so, it is important to note that collaborative leadership did make a significant contribution to growth in students' reading fluency and writing. Schools where teachers perceived strong collaborative leadership also displayed more positive perceptions of school climate, and more collaboration in both professional development and the delivery of reading instruction.

In terms of the reform effort, our results are mixed. With an average score of 4.2 on a 10-point scale, these schools clearly have a long way to go toward implementing all components of the reform. On a positive note, the one school that had made the most consistent gains in students' reading and writing growth was also the school in which teachers had made the most research-based shifts in the delivery of their classroom reading instruction.

Classroom-Level Findings

We believe that the most interesting findings in this study come from the observational data on classroom reading instruction, irrespective of school. The HLM analyses consistently found that higher-level questioning mattered: the more a teacher was coded as asking higher-level questions, the more that teacher's students' reading achievement improved. The teachers who were coded as asking more higher-level questions appear to be teachers who understand the importance of challenging their students' thinking about and comprehension of what they have read. There is little cause to celebrate, however: the positive impact of higher-level questioning is tempered by the low overall rates of occurrence of this practice across all of our K–6 observations. Furthermore, comprehension strategy instruction was seldom observed. Interestingly, the information in teachers' weekly logs corroborated this finding.

The findings on word skill work suggest that spending relatively large amounts of time on phonics instruction in grades 2–3 may not facilitate students' growth in reading fluency. This finding is compatible with the National Reading Panel's recommendation that phonics instruction should be concentrated in the earliest stages of schooling, mainly K and 1. Among older students, the practice of coaching in word recognition strategies during reading was found to be useful for students in grade 4–6. However, coaching in the application of phonics strategies is very different from explicit instruction focused on the letter-sound correspondences and rules, and is inherently more metacognitive and strategic in nature.

A negative relationship was also found between a highly teacher-directed stance towards reading instruction and reading growth for grades 2–6. This does not mean that teachers should never tell students information or engage them in recitation; it would be impossible to teach without doing so. However, it does appear that a heavy reliance on telling and recitation as teaching techniques is negatively related to children's reading growth. Excessive amounts of "telling," especially in situations where it would be possible to coach students to come up with their own responses, may rob children of the opportunity to take responsibility for their own skills and strategies. It may be useful to provide teachers with observational data on the frequency of telling and recitation in their literacy teaching, in order to help them shift away somewhat from a teacher-directed stance. This shift

would ideally lead to enhanced student performance. Over the coming years, as we provide teachers and schools with data on how teaching practices are tied to students' performance, we plan to investigate the degree to which classroom- and school-level teaching practices shift over time toward practices which have been identified as more effective for enhancing student achievement.

On the other hand, our classroom observations indicate that children in grade 1 showed less growth in writing when their teacher exhibited a strong student support stance (e.g., in cases where we saw relatively high levels of coding on the composite of modeling, coaching, and watching/giving feedback). This finding may be an anomaly, since it is contrary to other findings on the importance of scaffolding (Pressley et al., 2001; Taylor et al., 2000). Or it may be that children benefit from more teacher direction when learning to write in grade 1. Clearly, this is an area in need of further research.

The findings on grouping practices were mixed, and depended on the grade level in question. For reading comprehension in grade 1 and concepts of print in kindergarten, the coding of small-group instruction was positively related to students' reading growth. In grades 4–6, the coding of whole- or large-group instruction was positively related to students' reading growth.

Although our findings to this point are primarily limited to teachers, irrespective of school, we will continue to study schools to see if project participation might lead to building-level shifts. We will also continue to investigate the impact of teacher-level factors on students' reading and writing growth, as we have done in the current paper.

Limitations

One limitation of the study is that we were only able to investigate school- and classroom-level practices in five project schools during the first year, and eight schools in the second year. Only three of these eight schools have completed 2 full years in the project by this writing. Since change takes time, our ability to analyze the impact of the reform effort has thus far been limited by the small number of schools which have been in the project for at least 2 years.

In addition, classroom information was gathered from three one-hour observations per classroom, per year. This only gives us a snapshot of the reading instruction within these classrooms. The log information that was collected in Year 2 helps to provide a more complete picture of the reading activities occurring within certain classrooms, but unfortunately, log data were not available for the Year 1 schools.

Another limitation of the study is that some elements of the research design changed slightly from Year 1 to Year 2, due to the need to give participating teachers the best information available. In Year 2, we gave teachers detailed feedback on their observations, based on their requests. This feedback may have affected their teaching during the observations. In Year 2 we gave all schools a report summarizing the research findings from Year 1, because we

felt a moral obligation to give these schools our data as they became available. Fortunately, procedures between Years 2 and 3 of the project have not changed in any appreciable way.

Conclusions

The findings in this study complement previous research on effective schools and teachers. Schoolwide reading improvement should be the result of collaboration between the principal and teachers. Classroom literacy instruction needs to reflect best practices as identified in the research. In addition to considering *what* teachers teach, the current study's findings at the classroom level (which corroborate earlier research) suggest that it is important to consider *how* they teach, when seeking to make changes in reading instruction and significantly improve students' reading achievement. Based on the results of this study, the students whose teachers engage in higher-level questioning about the stories that they read show more growth in reading and writing during the year. Findings also suggest that a heavy reliance on telling and recitation, indicative of a strong teacher-directed stance, is not a very effective teaching strategy for enhancing students' reading growth.

Unfortunately, this particular study was not able to advance our knowledge about practices which involve parents in their children's learning. However, a considerable body of research already supports the notion that successful schools have found exemplary ways to involve parents as partners (Charles A. Dana Center, 1999; Designs for Change, 1998; Lein et al., 1997; Puma et al., 1997; Taylor et al., 2000). As we continue with this project, we hope to shed light on successful practices in this important area of school reform.

The improvement of our children's reading achievement is currently a major national goal (Bush, 2001). Schools know that there is a wealth of information available to help them move toward this goal, but the most relevant information is not always available in a format that helps schools take action. In the face of increasing pressure on schools and districts to adopt external programs, we remain optimistic about approaches that are locally developed and home-grown—as long as they are driven by the best research available on reading pedagogy and school change, and as long as they are enacted within a framework that features teacher involvement in and ownership of the change process. This approach, we think, will enable educators to create the knowledge base and sustain the commitment that are necessary in order to meet the ambitious goals we have set for ourselves as a nation.

References

Brophy, J. (1973). Stability of teacher effectiveness. *American Educational Research Journal*, 10, 245–252.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park, NJ: Sage.

Bush, G. W. (2001). *No child left behind.* Washington, DC: Office of the President of the United States.

Charles A. Dana Center, University of Texas at Austin. (1999) *Hope for urban education: A study of nine high-performing, high-poverty urban elementary schools.* Washington, DC: U.S. Department of Education, Planning and Evaluation Service.

Colt, J. (1997). *A scoring rubric for children's story retelling.* Longmont, CO: St. Vrain Valley School District.

Deno, S. (1985). Curriculum-based measurement: the emerging alternative. *Exceptional Children, 52,* 219–232.

Designs for Change. (1998). *Practices of schools with substantially improved reading achievement* [On-line]. Available: www.dfc1.org/summary/report.htm. Chicago: Chicago Public Schools.

Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly, 20,* 347–368.

Dunkin, M., & Biddle, B. (1974). *The study of teaching.* New York: Holt, Rinehart, & Winston.

Flanders, N. (1970). *Analyzing teacher behavior.* Reading MA: Addison-Wesley.

Fullan, M. (2000). *Change forces: The sequel.* London: Falmer.

Fullan, M., & Hargreaves, A. (1996). *What's worth fighting for in your school.* New York: Teachers College Press.

Greenwood, C. R., Carta, J. J., Kamps, D., & Delquadri, J. (1995). *Ecobehavioral assessment systems software (EBHASS) practitioner's manual, version 3.0.* Kansas City, KS: University of Kansas, Juniper Garden's Children's Project.

Herman, P. (1999). *An educator's guide to school-wide reform.* Washington, DC: American Institutes for Research.

Hoffman, J. V. (1991). Teacher and school effects in learning to read. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research, Vol. II* (pp. 911–950). New York: Longman.

Johns, J. (1997). *Basic reading inventory* (7th ed.). Dubuque, IA: Kendall Hunt.

Knapp, M. S. (1995). *Teaching for meaning in high-poverty classrooms.* New York: Teachers College Press.

Kreft, I., & De Leeuw, J (1998). *Introducing multilevel modeling.* London: Sage.

Lein, L., Johnson, J. F., & Ragland, M. (1997). *Successful Texas schoolwide programs: Research study results.* Austin, TX: Charles A. Dana Center, University of Texas at Austin.

Louis, K. S., & Kruse, S. (1995). *Professionalism and community in schools.* Thousand Oaks, CA: Corwin.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie Reading Tests* (4th ed.). Itasca, IL: Riverside.

Michigan Literacy Progress Profile 2000. (1998). Lansing, MI: Michigan Department of Education.

National Reading Panel (2000). *Report of the National Reading Panel.* Washington, DC: National Institute for Child Health and Human Development.

Pikulski, J. (1996). *The emergent literacy survey.* Boston: Houghton Mifflin.

Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching.* New York: Guilford Press.

Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., Baker, K., Brooks, G., Cronin, J., Nelson, E., & Woo, D. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading,* 5, 35–58.

Puma, M. J., Karweit, N., Price, C., Ricciuiti, A., Thompson, W., & Vaden-Kiernan, M. (1997). *Prospects: Final report on student outcomes.* Washington, DC: U.S. Department of Education, Planning and Evaluation Services.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2000). *HLM: Hierarchical linear and nonlinear modeling* (version 5) [Computer software]. Chicago: Scientific Software International.

Richardson, V., & Placier, P. (in press). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.). Washington, DC: American Educational Research Association.

Roehler, L. R., & Duffy, G. G. (1984). Direct explanation of comprehension processes. In G. G. Duffy, L. R. Roehler, & J. Mason (Eds.), *Comprehension instruction: Perspectives and suggestions* (pp. 265–280). New York: Longman.

Scanlon, D. M., & Gelzheiser, L. M. (1992). *Classroom observation manual.* Unpublished manuscript, University of Albany, State University of New York: Child Research and Study Center.

Soar, R. S., & Soar, R. M. (1979). Emotional climate and management. In P. Peterson & H. Walberg (Eds.), *Research on teaching: Concepts, findings, and implications*. Berkeley, CA: McCutchan.

Stallings, J., & Kaskowitz, D. (1974). *Follow through classroom observation evaluation* 1972–73 (SRI Project URU-7370). Stanford, CA: Stanford Research Institute.

Taylor, B. M. (1991). *A test of phonemic awareness for classroom use*. Minneapolis, MN: University of Minnesota.

Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary grade reading instruction in low-income schools. *Elementary School Journal*, 101, 121–166.

Taylor, B. M., Pressley, M. P., & Pearson, P. D. (2002). Research-supported characteristics of teachers and schools that promote reading achievement. In B. M. Taylor & P. D. Pearson (Eds.), *Teaching reading: Effective schools, accomplished teachers* (pp. 361–374). Mahwah, NJ: Erlbaum.

Venezky, R. L., & Winfield, L. (1979). *Schools that succeed beyond expectations in teaching reading* (Technical Report No. 1). Newark, DE: Department of Educational Studies, University of Delaware.

Weber, G. (1971). *Inner city children can be taught to read: Four successful schools* (CGE Occasional Papers No. 18). Washington, DC: Council for Basic Education. (ERIC Document Reproduction Service No. Ed 057 125)

Wilder, G. (1977). Five exemplary reading programs. In J. T. Guthrie (Ed.), *Cognition, curriculum, and comprehension* (pp. 57–68). Newark, DE: International Reading Association.

Ysseldyke, J., & Christenson, S. (1993–96). *TIES: The instructional environment system—II*. Longmont, CO: Sopris West.

Appendix A

RUBRIC FOR RATING INTERVIEW RESPONSES

0_____1_____2_____3

Low High

A. Building Collaboration (perception):

0—Teachers work in isolation or talk only at grade level, some sense of negative climate.

1—Only or mostly grade level talk, ambivalent climate, nothing mentioned about collaboration or a learning community, or it is mentioned only in passing.

2—Some talk across grades, but not a great deal, collaboration is mentioned but not stressed, teachers provide specific examples of how they are collaborating within their building, some sense of positive climate.

3—Cross-grade talk, collaboration on delivery of reading program, on professional development, collaborative learning community, positive climate.

B. Links to parents (school's efforts to reach out to parents):

0—Teachers expressed considerable dissatisfaction with parental involvement and little or nothing is being done by the school to facilitate a link with students' home environments.

1—Very little mentioned about parents, or teachers expressed dissatisfaction with parental involvement.

2—Some teachers actively pursue parental involvement in the classroom, mention that parents participate in opportunities offered at school (i.e., library reading program, parent center, site council, school meetings).

3—Includes those activities listed in Medium rating, but also includes a schoolwide focus, with teachers conducting phone or written surveys, interviews or focus groups to find out parents' concerns, teachers and/or principal calling home at least once a month with good news, as well as to discuss concerns, teachers sending home a newsletter or personal note at least once a week, anything else that the school does to invite parents in as partners.

C. Instructional reflection and change:

0—Little or no reflection on instructional practice by the individual classroom teachers, some talk between individual teachers about what is working.

1—Teachers talk and share ideas with each other about what is working in their classrooms during formal meeting times (i.e., grade level meetings)

2—Teachers talk and share ideas with each other in study groups. They may examine student work, reflect on their own instructional practice, and read current research on best practices, but most of their discussions focus on sharing what they do in their own classrooms.

3—Teachers indicate they are intentionally reflecting on their practice and are seriously working with others to improve their practice (i.e., study groups with action plans, grade level meetings to improve instruction), discussion within groups is informed by research on best practices and student assessment data

D. Views of professional development:

0—Teachers express dissatisfaction with the quality and quantity of professional development opportunities.

1—Teachers just mention professional development opportunities.

2—Teachers mention professional development opportunities and discuss what they have learned from district workshops, research (CIERA web site, journal articles, etc.) with other staff, there is some sense that teachers are trying to implement new ideas.

3—Professional development is ongoing, teachers have time to discuss, share, reflect on their practice, engage in professional development together across the building = collaborative learning community.

E. Leadership:

0—Teachers express dissatisfaction with their schools and the schools' administration.

1—Teachers express dissatisfaction with their school or may be detached from the problems of their school without taking responsibility for implementing change, teachers express low to moderate satisfaction with the school administration.

2—Some teachers assume instructional leadership in the school, teachers express moderate to high satisfaction with school administration.

3—Includes those activities listed in medium rating, as well as the following: principal or administrative staff are strong leaders who also get teachers involved in leadership, time is provided for teachers to operate as a collaborative learning community, leadership helps the school use data to reflect on where they are and where they want to be (not just student assessment data, but current research on best practices), teachers express high satisfaction with school administration.

F. Assessment:

0—Assessment is not used to inform instruction, teachers may feel pressured to "teach to the test."

1—Some teachers are using student assessment data to inform their own instructional decisions, but there is no school-wide alignment between assessments and the curriculum.

2—School has worked together to align assessment with curriculum, and is concerned about building-level assessment as well as state/district mandated assessments.

3—Includes the activities in medium rating, as well as the following: school uses assessment data to make changes in instruction, to change aspects of the reading program that don't appear to be working.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").